Unravelling the molecular pathogenesis of Myelodysplastic Syndrome Patients with Ringed Sideroblasts

Mian, Syed Asif

Awarding institution:
King's College London

Unravelling the molecular pathogenesis of Myelodysplastic Syndrome Patients with Ringed Sideroblasts

Syed Asif Mian

Thesis submitted for the Degree of Doctor of Philosophy

King's College London

October 2015

Declaration of Authorship

I hereby declare that I alone composed this thesis and that work presented here is my own, except where stated otherwise.

October 2015

Abstract

Myelodysplastic syndromes (MDS) are a collection of clonal hematopoietic stem cell disorders with diverse phenotypes, characterized mainly by ineffective haematopoiesis. Refractory anaemia with ringed sideroblasts (RARS) is an acquired form of MDS with accumulation of mitochondrial iron that gives rise to ringed sideroblasts. In this study, whole-exome sequencing performed in 12 patients all of whom had >25% ringed sideroblasts identified mutations in one particular gene, splicing factor 3b subunit 1 (*SF3B1*), a component of the major and minor spliceosomes, in 11/12 cases. No *SF3B1* mutations were observed in a single case of congenital sideroblastic anaemia, however, a constitutional mutation in *ALAS1* gene was confirmed in this patient.

Subsequently, amplicon sequencing using a 22 myeloid-gene panel in 154 MDS patients revealed 76% (n=117) of the patients had mutations in at least one of the genes, with 38% (n=59) having splicing gene mutations and 49% (n=75) patients harbouring more than one gene mutation. Interestingly, single and specific epigenetic modifier mutations tended to coexist with *SF3B1* and *SRSF2* mutations ($p<0.03$). Moreover, patients with splicing factor mutations alone had a better overall survival than those with coexisting additional mutations.

Next, combination of techniques including xenotransplantation and single cell clonogenic assays were used to study the characteristics of *SF3B1* mutant cells. Mutational analysis of hematopoietic stem/progenitor cells showed that *SF3B1* mutations arise in phenotypically defined $CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$ stem cells, therefore revealing the existence of MDS-initiating stem cells that propagate mutations to myeloid progenitors only. A persistent engraftment restricted to myeloid lineage was observed which was initiated only via the xenotransplantation of HSCs. Genetically diverse evolving subclones of mutant *SF3B1* were detected in mice, indicating a branching multi-clonal as well as ancestral evolutionary paradigm. Subclonal evolution in mice was also mirrored in the clinical evolution in patients. Sequential sample analysis uncovered a clonal evolution in hematopoietic progeny and selection of the malignant driving clone leading to AML transformation.

Acknowledgements

Apart from one's own efforts, the success of any project depends largely on the encouragement and guidelines of many others. I would like to take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of my PhD.

First of all I would like to show my greatest appreciation and extend my heartfelt gratitude to my supervisors Prof. Ghulam J Mufti and Dr. Alexander Smith for giving me the opportunity to carry out this work which was funded by Leukaemia and Lymphoma Research. I can't say thank you enough for their tremendous support and help which I received whilst working on this project. Without their encouragement and guidance this project would not have been successful.

I would also like to thank all the patients who have donated their bone marrow tissues to King's College London tissue bank.

Also, I would like to acknowledge and extend my heartfelt gratitude to the people who made the completion of my project successful including Prof. Dominique Bonnet and Dr. Kevin Rouault.

I owe my most sincere gratitude to Dr. Aytug Kizilors not only for her scientific advice but also for her deep understanding and help during my difficult moments.

I would like to thank my colleagues in the department of Haematological medicine for their help and support when it was needed, but most of all for making me feel at home amongst them. It was a pleasure to work with them all.

I would also like to thank all my friends especially Selen Bozkurt for her support and encouragement over the years.

Finally, I am forever indebted to my family especially my parents for their understanding, encouragement, and endless love 'come rain or shine'. I owe them so much.

Table of Contents

List of Figures

List of Tables

List of Publications

Syed A. Mian, Kevin Rouault-Pierre, Alexander E. Smith, Thomas Seidl, Irene Pizzitola, Dr Aytug Kizilors, Austin G. Kulasekararaj, Dominique Bonnetand Ghulam J. Mufti, (2015) SF3B1 mutant MDS initiating-cells may arise from the hematopoietic stem cell compartment. Nature communications.

Syed A. Mian, Alexander E. Smith, Austin G. Kulasekararaj, Aytug Kizilors, Azim M. Mohamedali, Nicholas C. Lea, Konstantinos Mitsopoulos, Kevin Ford, Erick Nasser, Thomas Seidl and Ghulam J. Mufti. (2013) Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica, 98, 1058-1066.

Kulasekararaj AG, Smith AE, **Mian SA,** Mohamedali AM, Krishnamurthy P, Lea NC, Gäken J, Pennaneach C, Ireland R, Czepulkowski B, Pomplun S, Marsh JC, Mufti GJ. (2013) *TP53* mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. BJH, 160, 660-672.

List of Abstracts

Syed A. Mian, Kevin Rouault-Pierre, Alexander E. Smith, Thomas Seidl, Austin G. Kulasekararaj, Azim M. Mohamedali, Sneha Shinde, Dominique Bonnet and Ghulam J. Mufti. (2013) SF3B1 mutant clones from patients with refractory anaemia with ringed sideroblasts (RARS) originate from the early Haematopoietic stem cells and maintain their engraftment potential. The American Society of Haematology, New Orleans, 55 Annual Meeting (Oral Presentation).

Alexander E Smith, **Syed A Mian**, Austin G Kulasekararaj, Azim M Mohamedali and Ghulam J. Mufti (2011) Whole Exome Sequencing Reveals Acquired *SF3B1* Mutations Defining Patients with Acquired Idiopathic Sideroblastic Anaemia. The American Society of Haematology, San Diego, 53rd Annual Meeting (Poster Presentation).

Abbreviations

AML	Acute myeloid leukaemia
ALL	Acute lymphoblastic leukaemia
ALA	Aminolevulinic acid (ALA)
adRP	Autosomal dominant retinitis pigmentosum
BFU	Burst froming units
BPS	Branch point splice-site
BWA	Burrows-Wheelers aligner
CCDS	Consensus Coding Sequence Project)
CFU-GM	Colony forming units-Granulocyte/Macrophage
CFC	Colony forming cell
CMP	Common myeloid progenitor
CML	Chronic myeloid leukaemia
CMML	Chronic myelomonocytic leukaemia
CDR	Commonly deleted region
DLBCL	Diffuse large B-cell lymphoma
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EF	Enhancing fluid
EPO	Erythropoietin
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
FACS	Fluorescent activated cell sorting cell sorting
FADD	Fas-associated protein with death domain
FDR	False disvocery rate
FC	Ferrochelatase
FISH	Fluorescence in situ hybridisation
GMP	Granulocyte macrophage progenitor
GCSF	Granulocyte-colony stimulating factor
GEP	Gene expression profiling
HSCT	Haematopoietic stem cell transplant
HSPC	Haematopoietic stem/progenitor cell
HSC	Haematopoietic stem cells.
HEC	Human engrafted cells
IGV	Integrated genome viewer
ISE	Intronic splicing enhancer
IPSS	International prognostic scoring system
IPSSR	International prognostic scoring system revised
IRP	Iron responsive protein
ISC	Iron responsive elements (IREs iron–sulfur cluster
IVT	In-vitro transcription
LTC	Long-term tissue culture
LTC-IC	Long term culture-initiating cell
LOH	Loss of heterozygosity
LMPP	Lymphoid-primed multipotent progenitor
MAB	Mutant allele burden
MACS	Magnetic activated cell sorting
MDS	Myelodysplastic Syndrome
MDS-RS	Myelodysplastic Syndrome with ring sideroblast
MHC	Major Histocompatibility complex
MDS-RS	MDS patients with ring sideroblasts
MEP	Megakaryocyte/erythrocyte progenitor
MC	Metaphase cytogenetics

MNC	Mononuclear cells
MID	Multiple-identifiers
MPC	Magnetic Particle Concentrator
MLP	Multilymphoid progenitor
MPD	Myeloproliferative disorders
MDS	Myelodysplastic syndromes
MPP	Multipotent progenitor
MDS-U	Myelodysplastic syndrome unclassified
NPHP-RC	Nephronophthisis-related ciliopathies
NADP	Nicotinamide adenine dinucleotide phosphate
NOD	Non-obese diabetic
NTD	N-terminal domain
NPM	Nextera PCR master mix
PCR	Polymerase chain reaction
PBS	Phosphate buffered saline
PTP	Picotiter-plate
Pre-mRNA	Precursor messenger RNA
RNA	Ribonucleic acid
RTA	Real time analysis
RA	Refractory anaemia
RARS	Refractory anaemia with ringed sideroblasts
RCMD	Refractory cytopenia with multilineage dysplasia
RCMD-RS	Refractory cytopenia with multilineage dysplasia and ringed sideroblasts
RARS-T	Refractory anaemia with ring sideroblasts and thrombocythemia
RAEB-1	Refractory anaemia with excess blasts-1
RAEB-2	Refractory anaemia with excess blasts-2
tMDS	Therapy-related MDS
SGS	Second-generation sequencing
snRNP	Small nuclear ribonucleoproteins
SMA	Spinal muscular atrophy
TAC	Transcriptome analysis console
TD	Tagment DNA
TfR1	Transferrin receptor
TE	Tris-EDTA
TK	Tyrosine kinase
UPD	Uniparental disomy
WES	Whole-exome sequencing
WGA	Whole genome amplified
WGS	Whole-genome sequencing
WHO	World Health Organization
XLSA	X-linked sideroblastic anaemia

Chapter 1

General Introduction

1.1 Myelodysplastic syndromes

Myelodysplastic syndromes (MDS) constitute a group of clonal hematopoietic stem-cell disorders with diverse clinical phenotypes. These diseases are characterized by varying magnitude of ineffective haemopoiesis and bone marrow morphological dysplasia. The sequelae of MDS result from the peripheral blood cytopenias and includes combinations of anaemia, thrombocytopenia and/or neutropenia (Mufti et al., 2008). Approximately 30% of the MDS patients transform to acute myeloid leukaemia (AML). The incidence of MDS is 3-4 cases per 100,000, however it increases markedly with age, and the median age at diagnosis is 73 years. Men have a significantly higher incidence than women (4.5 versus 2.7 per 100,000 per year) (Mufti et al., 2008; Scott and Deeg, 2010) and the overall survival following a MDS diagnosis varies between few months to more than 10 years (Mufti et al., 2008).

Several risk factors have been implicated in the pathogenesis of MDS including age, alcohol, smoking, ionizing radiation, immunosuppressive therapy, viral infection, benzene and other environmental leukaemogens. However, these risk factors seen occasionally are associated with up to 30% of the patients, who are often described as having secondary MDS. Moreover, majority of the secondary MDS cases are therapy-related MDS (tMDS) who have been previously treated with chemotherapy and/or radiotherapy.

Much of the phenotypic heterogeneity in MDS is likely to be caused by a variety of genetic lesions that contribute to disease pathogenesis. Recently, a multistep sequential model for the development of MDS has been proposed. According to this model (Figure 1), the interplay of mutations in the stem cells, epigenetic alternations within the bone marrow cells and the abnormal bone marrow microenvironment may allow for a predominant dysplastic clone to be established. The genetic alterations which lead to MDS are due to the dysregulation of genes involved in cellular pathways involved in pivotal roles in normal haemopoietic cell development.

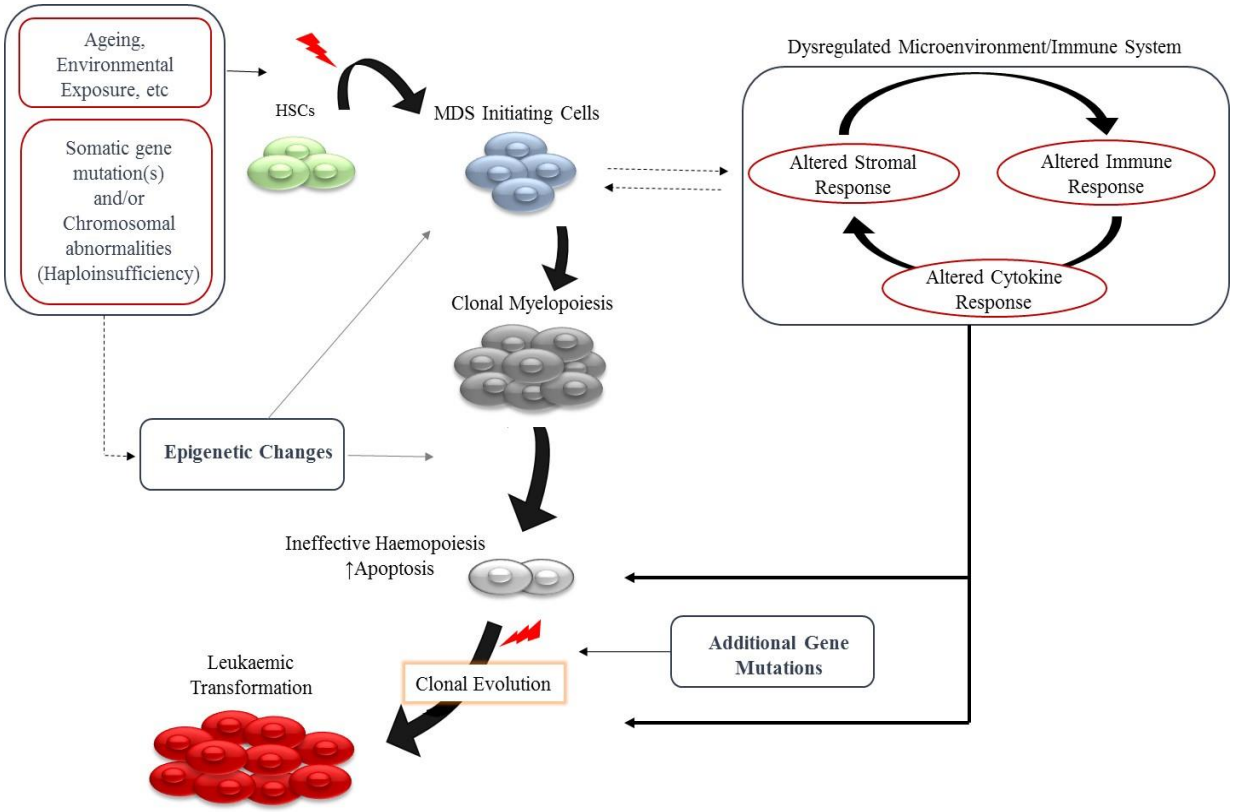


Figure 1: Pathogenesis of myelodysplastic syndromes

Age-induced genetic susceptibility and somatic mutations in haematopoietic stem cells lead to oligoclonal expansion of myelodysplastic stem cells, with defective differentiation, characterised by increased apoptosis of bone marrow progenitors. However, subsequent micro-environmental changes, immune dysregulation and epigenetic changes contribute to phenotypic diversity and susceptibility to leukemic transformation. MDS- Myelodysplastic syndromes, HSC- Haematopoietic stem cells.

1.1. MDS Classification

The phenotypic heterogeneity and the highly variable prognosis of MDS patients makes it difficult to classify the disease subtype, and predict the survival, and likelihood of leukaemic transformation. Therefore, over the past two decades several MDS classification and prognostic scoring systems have been proposed (Mufti et al., 2008; Scott and Deeg, 2010; Tefferi and Vardiman, 2009). Among the widely accepted systems, the French-American-British (FAB) classification widely modified by the World Health Organization (WHO) in 2008 is extensively used to classify MDS subtypes (Mufti et al., 2008). According to the WHO classification system for haematological cancers, patients with MDS are classified into 8 broad categories

(Table 1): Refractory anaemia (RA), Refractory anaemia with ringed sideroblasts (RARS), Refractory cytopenia with multilineage dysplasia (RCMD), Refractory cytopenia with multilineage dysplasia and ringed sideroblasts (RCMD-RS), Refractory anaemia with excess blasts-1 (RAEB-1), Refractory anaemia with excess blasts-2 (RAEB-2), MDS associated with isolated del(5q) and Myelodysplastic syndrome unclassified (MDS-U) (Scott and Deeg, 2010).

Table 1: Classification of MDS, according to World health Organization diagnostic criteria.

Disease	Blood findings	Bone marrow findings
Refractory anaemia (RA)	Anaemia No or rare blasts	Erythroid dysplasia only <5% blasts <15% ringed sideroblasts
Refractory anaemia with ringed sideroblasts (RARS)	Anaemia No blasts	Erythroid dysplasia only <5% blasts ≥15% ringed sideroblasts
Refractory cytopenia with multilineage dysplasia (RCMD)	Cytopenias (bicytopenia or pancytopenia) No or rare blasts <1 × 109/L monocytes	Dysplasia in ≥10% of cells in 2 or more myeloid cell lines <5% blasts in marrow <15% ringed sideroblasts
Refractory cytopenia with multilineage dysplasia and ringed sideroblasts (RCMD-RS)	Cytopenias (bicytopenia or pancytopenia) No or rare blasts <1 × 109/L monocytes	Dysplasia in ≥10% of cells in 2 or more myeloid cell lines ≥15% ringed sideroblasts <5% blasts
Refractory anaemia with excess blasts-1 (RAEB-1)	Cytopenias <5% blasts <1 × 109/L monocytes	Unilineage 5% to 9% blastsor multilineage dysplasia
Refractory anaemia with excess blasts-2 (RAEB-2)	Cytopenias 5% to 19% blasts Auer rods <1 × 109/L monocytes	Unilineage or multilineage dysplasia 10% to 19% blasts Auer rods
Myelodysplastic syndrome, unclassified (MDS-U)	Cytopenias No or rare blasts	Unilineage dysplasia in granulocytes or Megakaryocytes <5% blasts
MDS associated with isolated del(5q)	Anaemia <5% blasts Platelets normal or increased	Normal to increased megakaryocytes with hypolobated nuclei <5% blasts Isolated del(5q)

The WHO classification is supplemented by various prognostic scoring systems to predict the prognosis of the MDS patients. The most widely used prognostic scoring system used for stratification is the International Prognostic scoring System (IPSS) developed by Greenberg et al in 1997 (Greenberg et al., 1997). The IPSS is an important standard for assessing prognosis of primary untreated adult MDS patients and it groups patients into four risk categories based on the percentage of bone marrow blasts, number of cytopenias and karyotypic abnormalities detected by conventional cytogenetics (Table 2 and Table 3). Over the last few years, several MDS cooperating groups have revised the IPSS. This revised IPSS (IPSS-R) system identifies five prognostic categories. Bone marrow cytogenetics, bone marrow blast percentage and cytopenias remain the basis of the new system, however, a more precise refinement of some of these features (depth of cytopenias, splitting of bone marrow blasts <5% and more precise cytogenetic subgroups) demonstrated better separation of prognostic categories (Greenberg et al., 2012; Voso et al., 2013).

Table 2: The International prognostic scoring system (IPSS) values.

	0	0.5	1	1.5	2
BM Blasts (%)	<5%	5-10%	-	11-20%	21-30%
No. of Cytopenias⁺	0-1	2-3	-	-	-
Cytogenetics	Good: Normal, -Y, del(5q), del(20q)	Intermediate: Other abnormalities	Poor: Complex ≥3 abnormalities, chromosome 7 abnormalities	-	-
⁺ Platelet count <100 × 10^9 /L; Haemoglobin <100 g/L; Absolute neutrophil count <1.8 × 10^9 /L.					

Table 3: IPSS prognostic risk category clinical outcomes.

	Low	Intermediate 1	Intermediate 11	High
Risk score	0	0.5- 1.0	1.5-2.0	≥2· 5
Proportion of patients (%)	33%	38%	22%	7%
Median survival (years)	5.7	3.5	1.2	0.4
Time to AML evolution (years)	9.4	3.3	1.1	0.2

1.2 Spectrum of cytogenetic abnormalities in MDS

Unlike chronic myeloid leukaemia (CML) where the translocation t(9;22)(q34;q11) is the key single genetic hallmark of the disease, cytogenetic abnormalities in MDS are diverse. Over the years, several large cytogenetic studies have been published. *Greenberg et al* in 1997 published a multicentre study in which cytogenetic analysis was performed on 800 patients with primary MDS, which later formed an important component for the establishment of the IPSS (Greenberg et al., 1997). This was followed by Spanish study of 968 patients (Sole et al., 2005) and more recently German–Austrian study Group of 2,124 patients (Haase et al., 2007). These along with other studies have reported recurring chromosomal aberrations in up to 70% of patients where the type and complexity of the abnormalities correlates with prognosis, disease progression, leukaemic transformation and response to therapy (Table 4). Over 90% of tMDS patients have complex karyotype. A loss of genetic material in the form of deletions and monosomies are most frequent abnormalities. However, gain of genetic material with the appearance of total or partial trisomies is also seen in a smaller number of patients. Furthermore, gain and/or loss of genetic material as observed in unbalanced translocations are observed in patients with multiple cytogenetic abnormalities. This heterogeneity does not only hamper the prognostic classification but also has an impact on the delineation of the molecular background of cytogenetic aberrations. It is noteworthy, a normal karyotype is found in up to 40% of the MDS cases at the time of diagnosis (Bejar et al., 2011; Raza and Galili, 2012; Tefferi and Vardiman, 2009).

Table 4: Cytogenetic findings in patients with MDS, by their prognostic groupings.

Karyotype	MDS	Median Survival (Years)	Prognosis (IPSS)	Risk of Progression to AML
Del(20q)	2-5%	4.8	Good	Low
Del(5q)	15%	4.8	Poor	High
(The 5q syndrome)			Good	Low
Del7/(7q)	5-10%	2.7	Poor	High
17q	2-5%	2.7	Poor	High
Trisomy 8	5-8%	2.7	Medium	Intermediate
Monosomy Y	2-4%	2.7		Low
11q23	6%	5.4	Medium	Intermediate
Complex Karyotype	10-15%	0.7	Poor	High

1.2.1 Deletion 5q

Deletions of chromosome 5q (5q-) are the most frequent cytogenetic abnormalities in MDS deleted in approximately 15% of cases. The deletions vary in size between patients and are localized to two distinct commonly deleted regions (CDRs) on chromosome band 5q31. The proximal CDR at 5q31 is more centromeric and associated with bad prognosis, complex abnormalities and high-risk MDS as well as therapy-related MDS (tMDS). The more distal CDR in 5q33.1 is located in a more telomeric region and is associated with a clinical phenotype namely the 5q- syndrome. This is a good risk disease subgroup with a female predominance and, mainly characterized by severe macrocytic anaemia, normal/high platelet count, and a lower risk of leukaemic progression.

Over the past few years extensive sequencing of del(5q) MDS patients has not identified any mutations in genes present on the remaining intact allele of chromosome 5, therefore, suggesting that gene haplo-insufficiency might be the predisposition factor for the disease initiation. However, a recent study has reported mutations in 18% of the patients with del(5q) in *CSNK1A1* gene located within the CDR (Smith et

al., 2016) . In a separate study, *CSNK1A1* knockout mouse developed characteristics of bone marrow failure, therefore demonstrating the potential role of this gene in del(5q) MDS patients (Schneider et al., 2014). On the other hand, deletion of genes on 5q may not be involved in the development of MDS and additional aberrations might be required to provide additive effect to this del(5q) clone which gives rise to MDS.

Ebert et al performed a systematic functional RNA-mediated interference (RNAi)-based screening of the CDR of 5q in primary haematopoietic $CD34^+$ cells and identified *RPS14* as a critical gene for the erythoid/megakaryocytic phenotype of the 5q-syndrome (Ebert et al., 2008). The other genes implicated in the pathogenesis of del(5q) include *EGR1* and *CTNNA1*. However, deletions of 5q are not always limited to the CDRs and often encompass both of these regions and beyond. *NPM1* and *APC* are other genes that are located outside the CDRs but are lost with the deletions of 5q in MDS patients (Grisendi et al., 2005; Lane et al., 2010; Wang et al., 2010a). Moreover, other studies have demonstrated that haplo-insufficiency of two microRNAs located on chromosome 5q33, *miR-145* and *miR-146a*, can cause elevated platelet counts and may provide a selective advantage to the del(5q) clone (Starczynowski et al., 2010).

1.2.2 Chromosome 7 and 7q deletion

Monosomy 7 or interstitial loss of 7q is the second most frequent distinct chromosomal abnormality in MDS and is associated with poor prognosis. Chromosome 7 abnormalities, either in isolation or as part of a complex karyotype occurs in approximately 10% of the MDS patients (Haase et al., 2007). This frequency increases to 50% in tMDS patients who have a prior history of treatment with alkylating agents (Christiansen et al., 2004). Like del(5q) MDS, three common deleted regions on 7q have been identified including the band 7q22 and the more telomeric regions 7q31–32 and 7q36 (Asou et al., 2009; Dohner et al., 1998; Le Beau et al., 1996).

The underlying molecular mechanism driving the development of MDS in patients with chromosome 7 abnormalities are not well characterized. However, these interstitial as well as terminal deletions might be due to the cryptic unbalanced translocations. Previous studies have reported more than one different deletions in the same patients, either within the same copy of the chromosome 7 with retention of sequences between the deleted regions or within different mutant clones (Liang et al., 1998; Tosi et al.,

1999). These observations underline the chromosomal instability which makes regions within chromosome 7 prone to structural aberrations. *EZH2* gene localised to 7q36 is mutated in some patients with -7/del7q which alters the epigenetic state of haematopoietic stem cells (HSCs) by changes that occur in H3K27 trimethylation,a repressive mark in gene expression. However, conditional heterozygous deletion of a 2.5-Mb chromosomal 7 region (syntenic to human 7q22) in murine hematopoietic stem cells had no MDS phenotype indicating that this locus may not harbour a critical tumour suppressor gene (Wong et al., 2010). Moreover, gene expression analysis of $CD34^+$ cells from MDS patients with monosomy 7 has revealed over expression of *HOX9A*, *PRAME*, *BMI-1*, *PLAB* and the DNA repair gene *BRCA2* (Chen et al., 2004). In the same study, tumour suppressor gene *P21*, *GATA2* and *MAP* were found to be down regulated. It is noteworthy, additional abnormalities in monosomy 7 patients do not have such a profound impact on disease outcome unlike del(5q), since monosomy 7 even in isolation confers a significantly bad prognosis (Bejar et al., 2011; Haase et al., 2007).

1.2.3 Trisomy 8

Among the recurrent chromosomal abnormalities in MDS, trisomy 8 is the only chromosomal amplification observed in these patients. Trisomy 8 usually occurs in isolation in up to 8% of the MDS patients and is an intermediate-risk cytogenetic abnormality. Previous studies have reported over expression of anti-apoptotic genes in trisomy 8 cells and an expansion of Vβ-restricted $CD8^+$ T cells (Barrett and Sloand, 2009).

1.2.4 Other Cytogenetic Lesions

Several other recurrent cytogenetic abnormalities have been reported in MDS patients. These include del-Y, del20q, del3q26, del17p, isodicentric Xq13, and the t(6;9)(p23; q34) translocation that generates the *DEK-NUP214* fusion gene (Bejar et al., 2011). Patients with isolated del20q or -Y abnormalities are considered to be in the same favourable cytogenetic risk group as patients with a normal karyotype (Bejar et al., 2011). Translocation(s) and inversion(s) of chromosome 3q26 occur rarely in MDS and are associated with a poor prognosis (Bejar et al., 2011; Haase et al., 2007). Rest of the chromosomal aberration are rare in MDS and are grouped together in the intermediate-risk prognostic group (Bejar et al., 2011).

Although the mechanism underlying the generation of chromosomal abnormalities in MDS is not known; however, chromosomal shattering and re-joining of segments has been observed in many types of cancers, including leukaemia (Bassaganyas et al., 2013; Stephens et al., 2011). This process called chromothripsis, provides insights into the complex chromosomal abnormalities observed in MDS patients with abnormal karyotype and may be one of the mechanisms responsible for generating genetic diversity during MDS initiation and progression to AML. Chromothripsis (''chromo'' from chromosome; ''thripsis'' for shattering into pieces) is a term given to a process that involves massive *de novo* structural rearrangement in a one-step catastrophic genomic event. This process leads to the formation of tens to hundreds of locally clustered rearranged DNA fragments interspersed with widespread losses of sequences (often tens to hundreds) of the genome. In some instances, sequence rearrangements can lead to the formation of small circular DNA molecules (double-minute chromosomes), which may subsequently amplify if they harbour oncogenes. As a result of this, affected chromosomes show a characteristic pattern of copy-number "oscillations", in which only two or three copy-number states are observed along the chromosome in the context of a large number genomic rearrangements (Korbel and Campbell, 2013; Rausch et al., 2012; Stephens et al., 2011).

1.3 SNP Karyotyping

Recent application of sensitive technologies such as high-density single nucleotide polymorphism (SNP) arrays has led to the enhanced detection of smaller chromosomal aberrations, including micro deletions and discover uniparental disomy (UPD) with a neutral/overall loss of heterozygosity (LOH) occurring in up to 75% of MDS patients (Bejar et al., 2011). The first study where SNP arrays were used to screen MDS patients with normal karyotype detected frequent chromosomal aberrations with a high prevalence of UPD (Mohamedali et al., 2007) and this observation has been confirmed by other studies (Gondek et al., 2008; Heinrichs et al., 2009; Makishima et al., 2010b). A recent study used a combination approach to screen 430 patients with myeloid malignancies using SNP arrays as well as metaphase cytogenetics (MC). Notably, SNP karyotyping identified cryptic aberrations in low risk patients that were an independent predictor of outcome (overall survival, event-free survival and progression-free survival) in a dose-dependent manner (Tiu et

al., 2009). *Jerez et al* used SNP arrays to screen 1155 del(5q) patients and showed that the affected region on 5q determines clinical characteristics, with larger deletions involving the centromeric and telomeric segments of 5q associated with aggressive disease phenotype and additional chromosomal aberrations such as 17p (Jerez et al., 2012). In addition, large scale application of this technology in haematological malignancies has also led to the discovery of mutations in *TET2* (Langemeijer et al., 2009) (UPD 4q24), CBL (Dunbar et al., 2008) (UPD 11q) and EZH2 (Nikoloski et al., 2010) (UPD 7q) genes. Although balanced translocations which are commonly found in haematopoietic malignancies cannot be detected by SNP karyotyping but this technology complements and even outperforms conventional metaphase cytogenetic analysis, potentially allowing for the detection of more accurate genetic abnormalities in MDS.

1.4 Gene Mutations in MDS

The clinical heterogeneity observed in MDS has long been considered to be due to the accumulation of genetic lesions that culminate in the final disease phenotype. The genetic lesions include mutations that alter the protein structure (nonsense mutations, stop codons, splice-site mutations and/or insertions/deletions), or mutations in the enhancer and/or promoter regions of the genes that modify gene expression. Over the last decade, a number of genes are found to be mutated in MDS patients and these include known oncogenes and tumour suppressor genes. Recent technological advances in genomic assessments utilizing highly sensitive and scalable array- and next-generation sequencing-based methodologies have led to the identification of a wide variety of genes that are frequently mutated in MDS including genes involved in epigenetic regulation and cell signalling pathway/transcription regulation. Up to 60% of MDS patients have defects in one or more of these pathways, suggesting an underlying genomic instability or aberrant transcription regulation during the evolution of the disease. However, establishing the pathogenetic relevance of these individual genetic alterations has been a major challenge, primarily because of the multifaceted nature of the MDS phenotype and the broad spectrum of known molecular changes, affecting multiple and diverse pathways. Current IPSS used for MDS patient stratification doesn't not incorporate single or multiple gene mutations in the diagnostic classification or prognostic stratification systems, even though these mutations are

likely to be key drivers of clinical phenotypes and affect the overall survival and transformation to AML. However, as more mutations are identified and their consequences established, a profile of consequential genetic lesions has emerged that should allow genetic alterations to be incorporated into prognosis and risk stratification systems.

1.4.1 Epigenetic modifier gene mutations

The term ‘‘epigenetics’’ coined by Conrad Waddington is broadly used to describe chromatin-based events that regulate DNA-template processes such as DNA methylation, histone modification and nucleosome remodelling. Chromatin is the macromolecular complex acting as a scaffold for the packaging of human genome. The basic functional unit of chromatin is the nucleosome which contains 147 base pairs of DNA, which is wrapped around a core of histone octamer (2 each of Histones, H2A, H2B, H3, and H4). All these four histones have an amino-terminal tail that is lysine rich tail that protrude out of the nucleosome and is subject to posttranslational modifications (Luger et al., 1997). In addition, each nucleosome is separated by approximately 50 base pairs of DNA, which is packaged by the linker histone protein H1. Histones H2A and H2B are reported to play primarily a structural role, however, H3 and H4 are key integrators of a variety of signals that are involved in regulation of gene transcription. Histone modifications, especially the posttranslational modifications (such as acetylation, methylation, phosphorylation, ubiquitylation and sumoylation) occur within the “tails” or strings of amino acids that hang out of Histone proteins (H3 and H4). Studies have reported at least 16 classes of histone modifications (Table 5) (Dawson and Kouzarides, 2012). These post-translational modifications are important epigenetic mechanisms in controlling and maintaining chromatin structure and form a major category of epigenetic transcriptional control. These modifications can change chromatin structure by manipulating non-covalent interactions within and between nucleosomes which can either promote transcription, or result in gene silencing. They also serve as docking sites for specialized histone modifying proteins with unique domains that specifically recognize these modifications. These histone modifying protein are targeted to specific gene promoters by transcriptional activator/co-activator complexes in response to physiologic stimuli. Thus, histone modifications act as an “integral code” that combines signals for gene

activation/inactivation or silencing, therefore, enabling to predict the transcriptional activity of a given promoter by looking at the specific histone modifications.

The best studied histone modifications are methylation and acetylation of specific histone residues. For example, methylation of specific residues on H3 and H4 are associated with either activation of transcription (H3K4) or silencing (H3K9, H3K27). While as Acetylation of specific residues within these histone proteins is typically associated with active gene transcription. The different modifications can also interact, e.g. H3K4 methylation promotes H3K9 acetylation. However, it has been reported that several possible methylation states within the same residue can exist which can add further complexity to the system (Bannister and Kouzarides, 2011; Esteller, 2008; Galm et al., 2006). Mono and dimethylation within the H3K9 are associated with euchromatin silencing, and trimethylation of the same residue has been related with pericentric heterochromatin silencing. This H3K9 trimethylation leads to silencing via the recruitment of a number of proteins including HP1 proteins (Smallwood et al., 2007; Smith and Shilatifard, 2007; Tachibana et al., 2005). On the other hand, H3K27 trimethylation has been reported to occur early in X-chromosome inactivation, and leads to silencing through polycomb group (PCG) proteins (Okamoto et al., 2004).

Table 5: Known Chromatin Modifications and their Functions.

me1, monomethylation; me2, dimethylation; me3, trimethylation; me2s, symmetrical dimethylation; me2a, asymmetrical dimethylation; and Cit, citrulline. MBD, methyl-CpG-binding domain; PHD, plant homeodomain; MBT, malignant brain tumour domain; PWWP, proline-tryptophan-tryptophan-proline domain; BRCT, BRCA1 C terminus domain; UIM, ubiquitin interaction motif; IUIM, inverted ubiquitin interaction motif; SIM, sumo interaction motif; and PBZ, poly ADP-ribose binding zinc finger (Adapted from review on Cancer Epigenetics: From Mechanism to therapy (Dawson and Kouzarides, 2012)

Chromatin Modification	Nomenclature	Chromatin-Reader Motif	Attributed downstream Function
DNA Modifications			
5-methylcytosine	5mC	MBD domain	Transcription
5-hydroxymethylcytosine	5hmC	Unknown	Transcription
5-formylcytosine	5fC	Unknown	Unknown
5-carboxylcytosine	5caC	Unknown	Unknown
Histone Modifications			
Acetylation	K-ac	BromodomainTandem, PHD fingers	Transcription, repair, replication, and condensation
Methylation (lysine)	K-me1, K-me2, K-me3	Chromodomain, Tudor domain, MBT domain, PWWP domain, PHD fingers, WD40/b propeller	Transcription and Repair
Methylation (arginine)	R-me1, R-me2s, R-me2a	Tudor domain	Transcription
Phosphorylation (serine and threonine)	S-ph, T-ph	14-3-3, BRCT	Transcription, Repair, and Condensation
Phosphorylation (tyrosine)	Y-ph	SH2	Transcription and Repair
Ubiquitylation	K-ub	UIM, IUIM	Transcription and Repair
Sumoylation	K-su	SIM	Transcription and Repair
ADP ribosylation	E-ar	Macro domain, PBZ domain	Transcription and Repair
Deimination	R/Cit	Unknown	Transcription and Decondensation
Proline isomerisation	P-cis5 P-trans	Unknown	Transcription
Crotonylation	K-cr	Unknown	Transcription
Propionylation	K-pr	Unknown	Unknown
Butyrylation	K-bu	Unknown	Unknown
Formylation	K-fo	Unknown	Unknown
Hyroxylation	Y-oh	Unknown	Unknown
O-GlcNAcylation (serine and threonine)	S-GlcNAc, T-GlcNAc	Unknown	Transcription

In human genome, DNA methylation is another crucial epigenetic modification involved in regulating many cellular processes including embryonic development, transcription, chromatin structure, X-chromosome inactivation, genomic imprinting and chromosome stability. Studies have reported at least four different DNA modifications (Table 5). DNA methylation occurs only at cytosine bases that precede a guanosine in the DNA sequence (known as CpG dinucleotides) and is achieved by the covalent addition of a methyl group to the 5 position of a cytosine ring mediated by DNA methyltransferase (DNMTs) (Robertson, 2005). These CpG sites are asymmetrically distributed in the human genome, however there are small (0.5 kb to several kb) CpG-rich DNA regions known as CpG islands which are often located in the 5′ promoter end of all housekeeping genes and many tissue specific genes and with the 3′ end of some tissue-specific genes (Galm et al., 2006). Studies have shown that DNA methylation represses transcription directly by inhibiting the binding of specific transcription factors (TFs), and also indirectly, by recruiting methyl-CpG-binding proteins and their associated repressive chromatin remodelling activities (Robertson, 2005).

Alterations in DNA methylation have been implicated in the pathogenesis of MDS. This has been supported by observations in various studies showing that the global methylation pattern of cytosine residues in CpG dinucleotide sequences of promoter regions differ between MDS bone marrow cells and normal controls (Figueroa et al., 2009; Jiang et al., 2009; Shen et al., 2010). In addition, aberrant epigenetic modification of histone proteins have also been reported. There is now a growing genetic evidence that DNA methylation is important in MDS, because several genes that regulate cytosine methylation as well as histones are somatically mutated in MDS patients. These include DNA methyltransferase 3A (*DNMT3A*), Ten-eleven translocation 2 (*TET2*), isocitrate dehydrogenase 1 (*IDH1*) and *IDH2*, involved in Kerbs cycle and regulate the 5′ methylation of cytosines present in CpG dinucleotides in the genome. Two polycomb family member genes, additional sex-combs-like 1 (*ASXL1*) and enhancer of zeste homologue 2 (*EZH2*) that are involved in histone modifications are also mutated in MDS patients.

The 5-carbon of cytosine nucleotides are methylated (5mC) by a family of three active DNA methyltransferases (DNMTs). Among them *DNMT1* is a maintenance

methyltransferase that recognizes hemimethylated DNA generated during DNA replication. *DNMT3A* and *DNMT3B,* although also capable of methylating hemimethylated, are involved primarily in de novo methylation across the genome. *DNMT3A* is located on chromosome 2p23 and mutations of this gene were initially identified in normal karyotype AML. Subsequently, *DNMT3A* mutations were also found in a subset of MDS patients (<10%) with majority of mutations being heterozygous resulting in the amino acid substitution R882H. In addition, nonsense, compound heterozygous and frame shift mutations spanning across the whole protein are also found in MDS patients, suggesting that a loss-of-function mutations do exist. These mutations are found in all MDS subtypes and are associated with poor prognosis as well as rapid progression to AML. Homozygous R882H mutation reduces the methyltransferase activity of *DNMT3A in-vitro*, however, it has not been studied in combination with the wild-type allele. *Dnmt3a-/-* mouse models demonstrate an accumulation of phenotypically normal haematopoietic stem cells but with a decrease in their differentiation capacity, however, no difference was found in 5mC levels compared to wild type mice (Challen et al., 2012). Notably, distinct hypomethylation patterns were observed at pluripotency gene loci at *RUNX1* and *GATA3* with an increase in their gene expression. Interestingly, MDS patients with *DNMT3A* mutations have normal global 5mC levels, however, hypomethylation is seen in several individual gene loci which do not correlate with reduced gene expression profiles.

The TET protein family, initially found after the ten-eleven translocation (t(10;11)(q22;q23)) in human leukaemia's, turned out to be a key regulator for DNA demethylation (Figure 2). This group of proteins are 2-oxoglutarate and Fe(II)-dependent dioxygenases that have the capacity to converts 5-methyl cytosine (5mC) into 5-hydroxymethylcytosine (5hmC), which is then converted to unmodified cytosine through multiple mechanisms. TET proteins share the similar structural features with a carboxyl (C)-terminal that contains the catalytic double-stranded β-helix dioxygenase domain, which demonstrates oxidation activity against 5-methylcytosine (5-mC). This β-helix dioxygenase domain contains binding sites for α-ketoglutarate (α-KG) and Fe^{++} which are critical for their catalytic function. TET proteins also have a cysteine-rich domain (unknown function) that resides next to the double-stranded β-helix dioxygenase domain. The amino (N)-terminal region of *TET1* and *TET3* retains

an evolutionarily conserved CXXC-type domain that is critical for binding to unmethylated cytosine residues in DNA; however, this domain is absent in tet2 proteins.

TET2, localized on chromosome 4q24*, is* one of the most commonly mutated genes in MDS. Mutations of *TET2* are present in up to 20%-26% of the MDS patients, 44% of patients with MDS/MPN, and 11%-24% with sAML (Abdel-Wahab et al., 2009; Gondek et al., 2008; Mohamedali et al., 2007; Smith et al., 2010). *TET2* mutations (missense, nonsense and frameshift) result in loss of function are observed throughout the gene, with some clustering in the two conserved catalytic domains. Moreover, these mutations are associated with advanced age and normal karyotype (Smith et al., 2010). Unlike AML where *TET2* mutations are associated with poor prognosis (Abdel-Wahab et al., 2009; Metzeler et al., 2011), mutations in MDS do not appear to have prognostic significance (Smith et al., 2010).

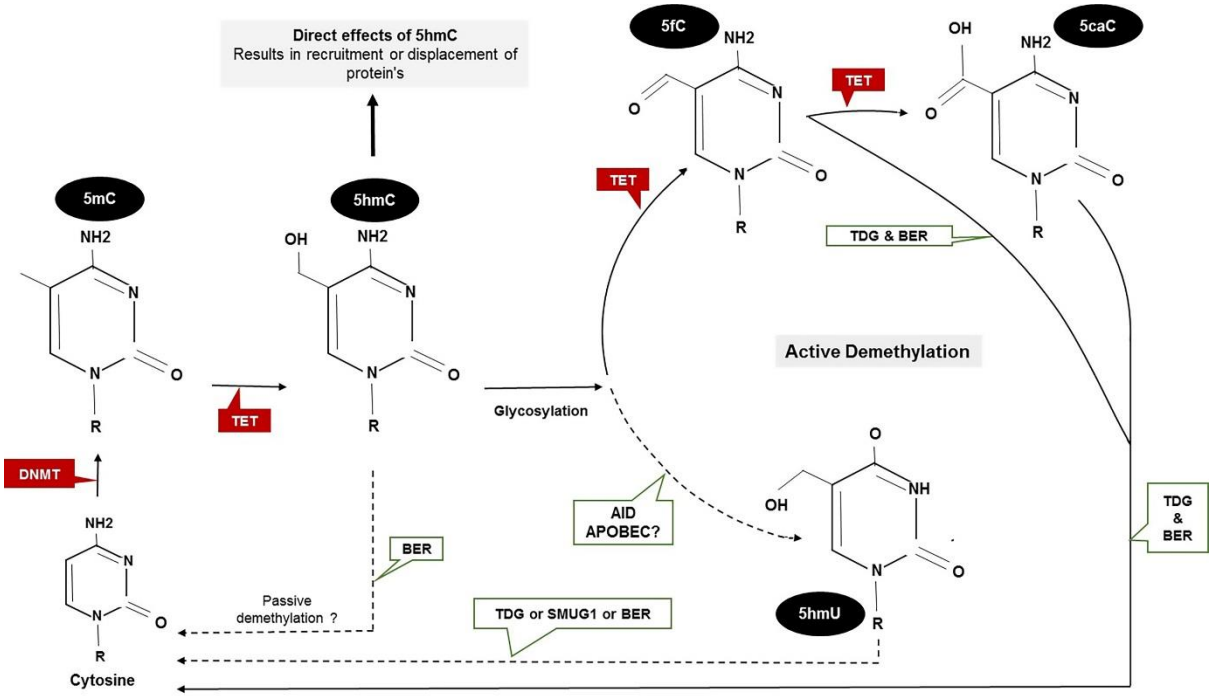


Figure 2: Schematics of TET-mediated cytosine demethylation.

Tet proteins sequentially oxidizes 5mC to 5hmC, 5fC and 5caC. 5caC and 5fC can be removed by thymine DNA glycosylase (TDG) and replaced by cytosine via BER. The other mechanisms of demethylation (shown by dotted line) shown in the figure are less well established, such as decarboxylation of 5caC, DNMT-mediated removal of the hydroxymethyl group of 5hmC and deamination of 5hmC (and 5mC) by the cytidine deaminases AID and APOBEC. Deamination of cytosine bases in DNA by AID enzymes produces uracil. It has been proposed that AID and APOBEC enzymes effect DNA demethylation by deaminating 5mC and 5hmC in DNA to produce thymine and 5hmU, respectively. As these are present as mismatched T:G and 5hmU:G nucleotides, they have been proposed to be excised by or TDG or SMUG1. 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine; BER, base excision repair; AID, activation-induced cytidine deaminase; 5-caC, 5-carboxylcytosine; DNMT, DNA methyltransferase; 5-fC, 5-formylcytosine; 5-hmC, 5-hydroxymethylcytosine; TDG, thymine DNA glycosylase; 5-hmU, 5-hydroxymethyluracil; 5-mC, 5-methylcytosine. SMUG1 - single-strand-selective monofunctional uracil DNA glycosylase; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide.

Previous studies have shown reduced levels of 5hmC in the bone marrow cells of patients with *TET2* mutations which is consistent with decreased catalytic activity of mutant *TET2* protein. This observation is also supported by mouse models, whereby, deletion of *TET2* leads to loss of 5hmC with a concomitant expansion of pluripotent

haematopoietic stem cells and a competitive growth advantage as well as a differentiation bias towards the myeloid lineage (Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011). *Moran-Crusio et al* have shown that reduced activity of *TET2* in haematopoietic stem and progenitor cells leads to clonal dominance, and enhances transformation to AML (Moran-Crusio et al., 2011). It is noteworthy that some of the MDS patients with wildtype *TET2* show decreased expression of tet2 in progenitors, erythroid precursors and granulocytes, implying alternative mechanism of protein deregulation (Langemeijer et al., 2009). However, xenotransplantation of $CD34^+$ cells from patients with *TET2* mutation showed expansion in NSG mice and *TET2* knockdown in cord blood haematopoietic progenitor cells led to aberrant myeloid differentiation (Delhommeau et al., 2009). It is important to note that these mutant cells did not show AML transformation in these assays, therefore implying *TET2*-mutated $CD34^+$ cells are not fully malignant, but rather behave as premalignant cells. In support of this observation, somatic mutations of *TET2* have recently been observed in healthy, aging individuals without an overt haematological disorder despite the presence of clonal haemopoiesis (Busque et al., 2012; Jacobs et al., 2012; Laurie et al., 2012).

Other recurrent genetic mutation in haematological malignancies include two key enzymes of tricarboxylic (TCA) cycle, *IDH1* and *IDH2*. Initially, these genes were found to be mutated in a high proportion of glioma and secondary glioblastoma (Yan et al., 2009). Subsequent studies using whole-genome sequencing (WGS) identified mutations of these genes in 16 of 188 patients with AML (Mardis et al., 2009) and 4% to 12% with MDS patients (Bejar et al., 2011; Kosmider et al., 2010; Thol et al., 2010). The mutations are heterozygous missense mutations of conserved codons (*IDH1*, R132; *IDH2*, R140/R172). No frame-shift or stop codon mutations were identified (Kosmider et al., 2010).

IDH1 and *IDH2* are homodimeric, nicotinamide adenine dinucleotide phosphate (NADP+) dependent enzymes that catalyse the conversion of isocitrate to α-ketoglutarate and reside in the cytoplasm and mitochondria, respectively (Mardis et al., 2009). Typical disease-associated mutations in these genes causes disruption of the isocitrate binding active site, resulting in the inability to make α-KG, instead reducing α-ketoglutarate to its homolog 2-hydroxyglutarate (2HG), which is a

metabolite normally found at very low concentrations in the normal cell (non-malignant cells) (Dang et al., 2009; Figueroa et al., 2010; Gross et al., 2010; Ward et al., 2010). The reduction in α-ketoglutarate results in DNA hypermethylation through the inhibition of multiple enzymes, including the tet proteins or jumonji D3, which require α-ketoglutarate and iron as a cofactor (Chowdhury et al., 2011; Xu et al., 2011). This report provides the biochemical basis for the mutual exclusivity of *IDH1/2* and *TET2* gene mutations observed in haematological malignancies (Metzeler et al., 2011; Paschka et al., 2010; Raza and Galili, 2012; Xu et al., 2011). Notably, mutant *IDH1/2* does not act in a dominant negative manner to its wild-type counterparts (Jin et al., 2011). Therefore, it has been proposed that disruption of *TET2* function by 2HG might be one of the mechanism for oncogenesis facilitated by mutant *IDH1/2*. Aberrant DNA hypomethylation has also been reported in patients with mutations of *IDH1/2* genes (Figueroa et al., 2010). Together, these findings suggest that instead of altering the expression of a single or a few specific genes, *IDH1* and *IDH2* mutations may change the expression of potentially large number of genes.

Polycomb proteins form chromatin-modifying complexes that are involved in transcriptional silencing. Two main components of Polycomb-repressive complexes (PRCs) are PRC1 and PRC2. The mechanisms by which PRCs repress target genes is poorly understood; however, it involves covalent modifications of histone tails catalysed by PRC1 (H2AK119 ubiquitination) and PRC2 (H3K27-me2,-me3), as well as direct chromatin modifications (Margueron and Reinberg, 2011; Simon and Kingston, 2009).

Mutations in two polycomb family member genes, *ASXL1* and *EZH2*, are relatively common in MDS. *ASXL1* located on chromosome 20, is mutated in 10-20% of MDS and 5-25% of AML patients (Boultwood et al., 2010; Gelsi-Boyer et al., 2009). *ASXL1* gene belongs to a highly conserved three-member family of enhancers of trithorax and polycomb proteins (ASXL-1, -2, -3) which are involved in the maintenance of activation as well as the silencing of development-related genes through chromatin remodelling. *ASXL1* contains the plant homeo domain (PHD) finger and nuclear receptor box domains, and functions as a ligand-dependent coactivator of the retinoic acid receptor (Fisher et al., 2006). Previous studies have shown that *ASXL1* mediates its effects

through direct interaction with the histone demethylase encoded by *LSD1 or* histone acyltransferase encoded by *SRC1* to control the chromatin state (Lee et al., 2010a). Mutations of *ASXL1* are mostly localized to exon-12 which is a glycine-rich region and, cause premature truncation of the protein with loss of C-terminal part of the protein (Bejar et al., 2011). *ASXL1* gene locus does not have a high frequency of chromosomal aberrations unlike other genes (such as in case of del 5q and 7q-) which are commonly mutated in MDS, and appears to be excluded from the common deleted region of chromosome 20q (Boultwood et al., 2010; Gelsi-Boyer et al., 2009).

The mechanism through which *ASXL1* mutations contribute to the pathogenesis of MDS is still unknown. Targeted disruption of *Asxl1* in mouse models have shown only mild haematopoietic abnormalities with reduced numbers of lymphocytes and modest splenomegaly; however, no evidence of myeloid dysplasia was observed (Fisher et al., 2010). Previous studies have shown that truncated *ASXL1* protein is unable to bind methylated histone lysines and is also not able to interact with other chromatin modifiers. This results in the loss of polycomb repressive complex (PRC) mediated H3K27 trimethylation and activation of a *HOXA* gene expression signature (Abdel-Wahab et al., 2012). This truncated protein with intact N-terminal motifs might be potentially generating a dominant negative protein that could inhibit its wild-type isoform as well as other members of the polycomb complex (Bejar et al., 2011). *In-vitro* studies in non-hematopoietic cells have suggested physical cooperativity of *ASXL1* with *HP1a* and *LSD1* to repress retinoic acid-receptor activity, and its interaction with peroxisome proliferator-activated receptor gamma (PPARg) to suppress lipogenesis (Cho et al., 2006; Lee et al., 2010a; Park et al., 2011). Interestingly, a recent study using gene expression profiling of *ASXL1* mutant patients have revealed significant deregulation of retinoic acid receptor activation pathways (Boultwood et al., 2010).

Another polycomb gene mutated in haematological malignancies is *EZH2,* a histone H1 and H3 methyltransferase gene that is a catalytic subunit of PRC2. Mutations of *EZH2* are present in approximately 6-8% of the MDS patients. *EZH2* alongside with other proteins (EED, SUZ12 and RBBP4) acts via a number of mechanisms to initiate polycomb-mediated gene repression. One of the mechanisms include direct inhibition of transcriptional machinery via RNA polymerase II and chromatin compaction,

therefore inhibiting the access and action of transcription factors. *EZH2* has also been reported to play a role in DNA methylation, by acting as a recruitment platform for DNA methyltransferases (Francis et al., 2004; Vire et al., 2006). Following trimethylation by the suppressor of variegation 3–9, the histone mark H3K27me3 functions as a docking spot to recruit polycomb complex PRC1, which maintains gene repression via ubiquitination of H2AK119 (Cao et al., 2002; Wang et al., 2004). In recent studies, *EZH2* has also been reported to regulate cell fate decisions by orchestrating expression of genes to control the balance between self-renewal and differentiation (Bracken et al., 2006). *EZH2* stabilizes the chromatin structure and maintains long-term self-renewal capacity of HSCs by turning off pro-differentiation genes and stimulating proliferation by increasing expression of cell cycle genes (Bracken et al., 2003; Ezhkova et al., 2009; Kamminga et al., 2006).

The occurrence of mutations within chromatin regulators such as *EZH2* has been linked to aberrant gene expression and genomic instability. *EZH2* missense mutations commonly occur within the C-terminus catalytic SET (suppressor of variegation, enhancer of zeste, trithorax) domain and cysteine-rich domain while truncation mutations have been reported throughout the gene (Nikoloski et al., 2010). This gene is located on chromosome 7 and mutations are associated with uniparental disomy (UPD) on 7q. *In-vitro* studies have suggested that the *EZH2* mutations found in MDS patients are predominantly loss of function (resulting in reduced H3K27 trimethylation) and may be associated with relatively poor survival (Ernst et al., 2010; Makishima et al., 2010a). A conditional Ezh2 'gain-of-function' mouse model demonstrated that induction of wildtype Ezh2 expression increases the repopulation capacity of HSCs (Herrera-Merchan et al., 2012). In another study, *Mll-AF9* transformed AML cells from Ezh2-deficient mice showed compromised cell cycle progression, although cells still maintained some proliferative capacity. In this model, Ezh2 loss inhibited leukemic progression by reactivating Ezh2-repressed genes involved in myeloid differentiation (such as Egr1*)* (Tanaka et al., 2012)*.* Another study has shown that loss of PRC2 function increases hematopoietic stem cell activity and expansion (Majewski et al., 2010), which may explain the role of *EZH2* mutations in MDS.

1.4.2 Transcription regulation/Cell signalling pathway gene mutations

The *TP53* gene, located on chromosome 17p, is a prototypical tumour suppressor gene that is frequently mutated in nearly all tumour types and is often associated with genomic instability. In MDS, *TP53* mutations occur predominantly in high risk disease subgroups, associated with complex karyotype, adverse prognosis, therapy-related disease and resistance to therapy. However, *TP53* mutations have also been reported in low risk MDS such 5q- syndrome (Jadersten et al., 2011). Alterations in the *TP53* gene occur in up to 15% of *de novo* MDS patients, however, the frequency is much higher in patients with prior exposure to alkylating agents and/or radiation (Christiansen et al., 2001; Pedersen-Bjergaard et al., 2006). In addition, a third to half of the MDS patients with complex karyotype have *TP53* mutations. Notably, *TP53* mutations are associated with poor prognosis in MDS, even after adjustment of other prognostic variables in IPSS, including complex karyotypes (Horiike et al., 2003; Kita-Sasai et al., 2001). Apart from its tumour suppressor activity, *TP53* gene is also a transcription factor that reacts to cellular stress by activating various defensive pathways that cause cell-cycle arrest, induce apoptosis or trigger DNA repair. Loss of *TP53* results in resistance to apoptosis and other cellular death signals. *Barlow et al in* a mouse model demonstrated that normally functional *TP53* is essential to initiate apoptosis promoted by haploinsifficiency of *RPS14* in del(5q) (Barlow et al., 2010). In this mice study, segmental haploidy CDR locus (5q- syndrome) residing in the interval between Cd74 and Nid67 resulted in features of 5q- syndrome (macrocytic anaemia, with prominent dysplasia within the erythroid lineage and monolobulated megakaryocytes). Among the 8 genes which reside within the Cd74-Nid67 deleted segment, the Rps14 gene was the strongest candidate associated with this phenotype (Barlow et al., 2010), due to its association with 5q- syndrome (Ebert et al., 2008). A high population of bone marrow cells in Cd74–Nid67 deleted mice expressed higher amounts of Tp53 protein as compared to controls. Notably, the deletion of *TP53* in Cd74–Nid67 deleted mice completely reversed the observed 5q- syndrome phenotype (Barlow et al., 2010). Therefore, *TP53* is stabilized as a result of haploinsufficiency of *RPS14* in del(5q) which causes disruption of ribosome assembly and accelerates degradation of *MDM2*.

Another frequently mutated gene in MDS is *RUNX1*, which is a member of the transcriptional core-binding factor gene family that was originally identified as a

translocation partner t(8;21) in AML(M2). Mutations in *RUNX1* gene are found in approximately 7%-15% of patients with *de novo* MDS and the frequency is much higher (up to 40%) in therapy-related MDS. The mutations are usually heterozygous missense and monoallelic (nonsense or frameshift), occur within the N-terminal while as few mutations have been reported in the C-terminal (trans-activation domain) of *RUNX1* (Christiansen et al., 2004; Harada and Harada, 2009). Studies thus far have not investigated exons 7/8 in detail which encode for most of the C-terminal region. Thus, C-terminal mutations within RUN*X1* may be underestimated.

The *RUNX1* contains a highly evolutionary conserved proximal Runt homology domain (RHD) which is essential for DNA binding, and a distal transactivation domain responsible for protein-protein interactions and recruitment of cofactors. Missense mutations *of RUNX1* are observed within RHD, whereas stop codon and frame shift mutations are present throughout the gene and almost always disrupt the transactivation domain (Harada et al., 2004). This distinction may be physiologically relevant since mutations in the RHD impairs DNA binding and therefore function in a dominant negative manner. These mutations are able to diminish Runx1 activity much more than loss-of-function (Harada and Harada, 2009; Matheny et al., 2007). Runx1 knockout mice are embryonically lethal with no evidence of definitive haemopoiesis (Dowdy et al., 2010; Okuda et al., 1996). However, heterozygous conditional deletion of this gene in the haematopoietic stem cells in mice lead to an increase in myeloid progenitors, lymphoid defects and ineffective megakaryopoiesis, but do not progress to AML (Growney et al., 2005). In contrast, mice overexpressing RHD mutant form shows increased bone marrow blasts and splenomegaly, and often die as a result of AML- like disease (Watanabe-Okochi et al., 2008).

Genes involved the RAS signalling pathway (*NRAS/KRAS*) are mutated in up to 15% of MDS patients, with *NRAS* being the common aberration (Haferlach et al., 2011). *RAS* genes are involved in the regulations of growth and differentiation of several cell types. *RAS* mutations cause constitutive activation of the *RAS* signalling pathway by increasing the intracellular levels of RAS GTP. This increase in RAS GTP leads to activation of the RAS/PI3K and the RAS/Raf/MEK signalling pathways through interaction with many effector molecules such as phosphoinositide-3-OH kinase, Raf proteins and Ral guanine nucleotide dissociation stimulator proteins (such as

RalGDs). Oncogenic *NRAS* or *KRAS* in mice is sufficient to induce AML or a myeloproliferative disorder that resembles chronic myelomonocytic leukaemia (CMML) (Parikh et al., 2006). This phenomenon has been shown to occur in hematopoietic stem cells and not in the common myeloid progenitor compartment (Wang et al., 2013).

C-CBL, located on chromosome 11 is a proto-oncogene that encodes a tyrosine kinase (TK)-associated ubiquitin ligase and negatively regulates signalling through these receptors by targeting them for degradation. It is mutated in approximately 5% of the MDS patients and most of these aberrations are missense mutations. Frame shift and early truncation mutations are rarely observed (Sanada et al., 2009). These *CBL* mutations occur in exons 8/9 and usually leave the rest of the gene intact, suggesting a gain-of-function/altered function effect of these mutations. *C-CBL* mutations have been shown to cause increased receptor TK levels and *STAT5* phosphorylation, which is believed to mediate the hypersensitivity of the mutated cells to cytokines and growth factors (Sargin et al., 2007; Saur et al., 2010). C-Cbl(-/-) haematopoietic stem/progenitor cells (HSPCs) show an increased sensitivity to a variety of cytokines, and transduction of *C-CBL* mutants into C-Cbl(-/-) HSPCs further augments their sensitivities to a broader spectrum of cytokines, including stem-cell factor, thrombopoietin, IL3 and FLT3 ligand (Sanada et al., 2009). Moreover, mouse cells deficient in CCBL show modest cytokine hypersensitivity (Rathinam et al., 2008). Activating mutations of *JAK2* were first described in myeloproliferative disorders (MPDs) and form part of the diagnostic criteria for polycythemia vera and are found in up to 90% patients with polycytemia rubra vera. *JAK2* V617F mutation is present in 5% of the MDS patients, however, the frequency is much higher (upto 50%) in a particular subtype of MDS patients i.e. refractory anaemia with ring sideroblasts and thrombocythemia (RARS-T) (Boissinot et al., 2006; Malcovati et al., 2009; Szpurka et al., 2006). The mechanistic role of this *JAK2* mutation in the RARS-T phenotype is unclear, and its contribution to prognosis is as yet not clearly defined.

Although RARS-T and MPD appears to have entirely different pathophysiological mechanisms, the existence of conditions with overlapping features is well recognised (Steensma and Tefferi, 2008; Wardrop and Steensma, 2009). RARS patients develop thrombocythemia only after the acquisition of *JAK2* V617F mutation (Malcovati et al.,

2009). The high rate of *JAK2* mutation in RARS-T patients points towards an unknown mechanism through which a cooperative advantage exists between those abnormalities responsible for the ring sideroblast phenotype and the constitutive activation of *JAK2*.

1.5 Spliceosome

Majority of the eukaryotic genes consists of exons (coding sequence) and introns (non-coding sequence), and are initially expressed as precursor messenger RNA (Pre-mRNA) that is converted to mRNA by a process called RNA splicing. This highly dynamic process is a crucial, ubiquitous and a complex step in gene expression in which introns are excised from pre-mRNA, and coding sequences (i.e. exons) are ligated together (Krummel et al., 2010; Ritchie et al., 2009; Wahl et al., 2009; Will and Luhrmann, 2011).

Pre-mRNA introns in protein-encoding genes are precisely excised by a complex called the Spliceosome that requires hydrolysis of a large quantity of ATP. The spliceosome is approximately 4.8 MDa protein composed of 5 uridine-rich small nuclear RNAs (Sn RNAs) and up to 300 other distinct proteins. This large protein-RNA complex is assembled on pre-mRNA by the stepwise association of several small nuclear ribonucleoproteins (snRNPs) and non-snRNP factors, recognizing (1) the junction between 5/ exon and intron (5' splice-site), (2) a conserved sequence within the intron called branch point-site (BPS) located ~15–50 nucleotides upstream of the 3'ss, and (3) the junction between the intron and 3' exon (3' splice-site) (Krummel et al., 2010; Ritchie et al., 2009; Wahl et al., 2009). Dinucleotides that form the intron/exon boundaries at 5' and 3' sites act as donor and acceptor splice sites (ss) for intron removal. In addition to being controlled by the primary splicing signals located at these exon/intron junctions, splice site choice is modulated by various *cis*-acting regulatory RNA elements that serve as either splicing enhancers or silencers. These *cis*-acting regulatory elements embedded in the pre-mRNAs recruit trans-acting splicing factors (bind to the cis-acting elements) that ultimately control assembly of other proteins involved in the splicing process. These cis-regulatory or splicing regulatory elements (SREs) are classified as exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) or intronic splicing silencers (ISSs) depending upon their location and activities. A well-studied example

is the oligo-G tracts which are known to enhance splicing from intronic locations by recruiting hnRNP H, but the same element when located in exons can inhibit splicing (Caputi and Zahler, 2001; Chen et al., 1999; Chou et al., 1999; McCullough and Berget, 1997).

Two types of spliceosomes coexist in most eukaryotic cells: the U2-dependent spliceosome (major spliceosome), which catalyses the splicing of U2-type introns with the canonical dinucleotide GU–AG, and the rare U12-dependent spliceosome (minor spliceosome), which splices the rare U12-type class of introns with canonical dinucleotide AT–AC and GT–AC (Will and Luhrmann, 2011). The two spliceosomes share most of the snRNP proteins which are involved in splicing reactions; however, the minor spliceosome is 100-fold less abundant compared to major spliceosome complex. U1, U2, U4, U5 and U6 are core snRNPs in major spliceosome complex, which along with other non-snRNP factors form pre-complex E, Complex-A, Complex-B, Complex-B* and Complex-C (Figure 3), respectively. However, in addition to the common U5 snRNP, the minor spliceosome contains other distinct snRNPs (U11, U12, U4atac, and U6atac), which are functionally similar to the U1, U2, U4 and U6 snRNPs of the major spliceosome complex (Konig et al., 2007; Will and Luhrmann, 2011).

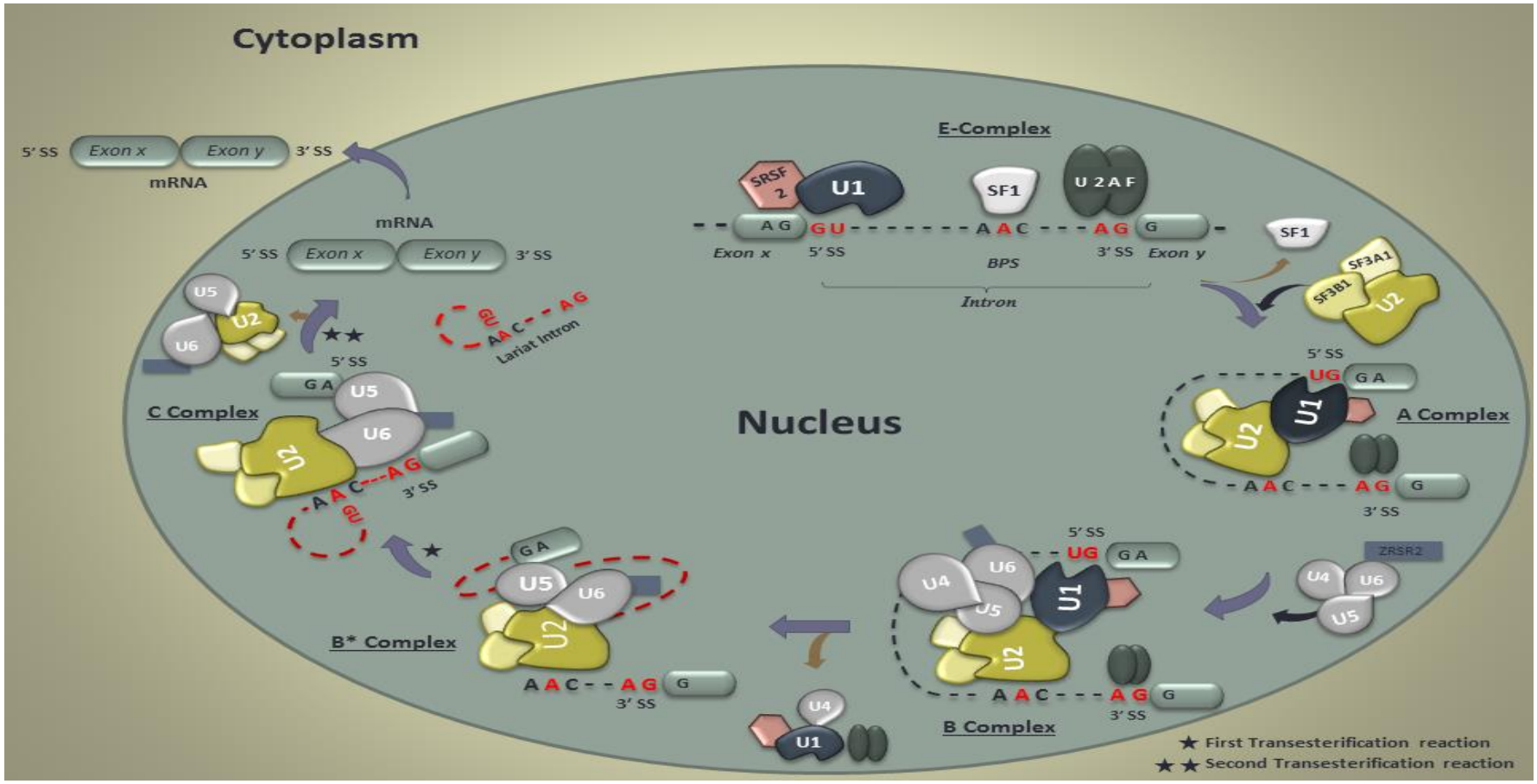


Figure 3: Schematic diagram showing the mRNA splicing pathway.

Major Spliceosome proteins catalyse RNA splicing, which leads to the ligation of flanking exonic regions. The GU dinucleotides at the 5′ splice-site, branch point site (BPS) and AG dinucleotide at the 3′ splice-site are specific recognition sites for splicing. Various spliceosome components including U1, U2, U4/U5/U6 snRNPs and other additional non-snRNPs assemble in a step-wise manner to form 5 different complexes (E, A, B, B* and C). The rectangles and broken lines represent exonic and intronic regions, respectively.

Although generally considered as a post-transcriptional process, mRNA splicing can also occur as a co-transcriptional process in the nucleus. Most of the components of the major spliceosomes are partly synthesized in the cytoplasm, however, some mature forms are mainly localized in the nucleus. In major spliceosome, the process of spliceosome assembly starts by U1 snRNP recognizing/binding at the 5'ss through base-paring interactions of the 5'end of the U1 snRNP to the 5'ss of the target intron, followed by the binding of some non-snRNP proteins to the BPS and 3' splice-site (such as U2AF) to form a pre-complex E. The interaction between the U1 snRNP and 5'ss in pre-complex E is ATP independent and fairly weak; however, it is stabilized by the binding of other factors at 5'ss such as SR proteins such as *SRSF2* and the cap-binding complex. This pre-complex E enables the pre-mRNA to commit for splicing. Following on, associated factors such as splicing factor 1 (SF1) and U2 auxiliary factors (U2AFs), which are also components of complex E bind to the 3´ss and BPS within the pre-mRNA (Matera and Wang, 2014; van der Feltz et al., 2012; Wahl et al., 2009).

Subsequently, U2 snRNP binds to the BPS in an ATP-dependent process catalysed by the DExD/H helicases pre-mRNA-processing 5 (Prp5) and Sub2, through base paring interaction with the target pre-mRNAs BPS, leading to the formation of complex-A. This base-paring is stabilized by heteromeric protein complexes SF3a and SF3b which are both part of the U2 snRNP. U1 snRNP intracts with the U2 snRNP which brings the 5'ss, BPS and 3'ss into close proximity. Recruitment of preassembled U4/U6-U5 tri-snRNPs follows along with non-snRNP such as *ZRSR2*, forming complex-B in a reaction catalysed by the DExD/H helicase Prp28. The resulting complex B goes through a series of conformational and compositional remodelling to form a catalytically active complex B (Complex-B*) and commits it for the first step of splicing reaction. Multiple RNA helicases (Brr2, 114 kDa U5 small nuclear ribonucleoprotein component and Prp2) are required during this process (Matera and Wang, 2014). Following on, some structural/compositional rearrangements occur within the complex which causes destabilization of U1 and U4 snRNPs, with U1 being displaced by U6 to form complex-C that catalyzes the final splicing reaction (Figure 3). This first splicing reaction yields the free exon (exon X) and the intron–exon (exon Y) lariat intermediate. Complex C undergoes further ATP-dependent rearrangements

before carrying out the second catalytic step of splicing reaction (dependent on Prp8, Prp16 and synthetic lethal with U5 snRNA 7) and this results in a post-spliceosomal complex that contains spliced exons (exon X and exon Y) and the lariat intron. Subsequently, the spliceosome dissociates, releasing the spliced mRNA into the cytoplasm for protein translation and this process is catalysed by the DExD/H helicase Prp22. In addition, U2, U5, and U6 snRNPs are released from the complex which are later recycled for further rounds of splicing (Faustino and Cooper, 2003; Krummel et al., 2010; van der Feltz et al., 2012; Wahl et al., 2009; Will and Luhrmann, 2011). This disassembly of the post-catalytic spliceosome mediated by several RNA helicases (such as Brr2, Snu114, Prp22 and Prp43) is ATP-dependent (Matera and Wang, 2014; Wahl et al., 2009).

In higher eukaryotes, most genes undergo differential splicing to give rise to large number of alternative mRNA isoforms, which generates functionally diverse protein isoforms with altered regulatory regions. These alternative splicing events are highly regulated, cell type-specific and controlled by numerous *cis*-acting regulatory RNA elements. Alternative splicing adjusts gene product in accordance with the physiological requirements of a cell, and more than 90% of genes with introns undergo alternative splicing, probably influencing all aspects of cell biology (Kalsotra and Cooper, 2011). Human introns range from few hundreds to thousands of kilobases in length (~5 kb on average). These large intronic regions contain “decoy” splice-sites (i.e. sequences that have a similar degree of consensus matching to actual splice-sites). A pair of these artificial “decoy” splice sites often leads to the formation of pseudo-exons that resembles authentic exons in terms of splice-site junction sequence but are very rarely, if ever, spliced. Thus, despite the prevalent decoy splice-sites, the splicing process occurs with high fidelity, implying that additional sequence factors aside from core splicing signals contribute to definition of exon/intron junctions (Sun and Chasin, 2000). Many splicing factors function as auxiliary proteins in the spliceosome pathway and interact with its core components (such as U1, U2, U3, U4/U5/U6) to regulate splicing. A well-studied example is the splicing factor RNA-binding motif 5 (RBM5) which interacts with a U2 snRNP component (U2AF65) and inhibits the transition from an exon definition to an intron definition complex (Bonnal et al., 2008). Therefore, alternative splicing can contribute to the genomic complexity of

the cell as well as the selective pressure that may occur in cancer genome (Hahn and Scott, 2012; Padgett, 2012; Ward and Cooper, 2010).

Splicing of most of the introns happens co-transcriptionally, therefore, alternative splicing can effect factors that are involved in controlling transcription initiation and elongation processes and vice-versa. For example, studies have shown that splicing events can be affected by the rate at which the transcription elongation occurs with slow elongation rates generally promoting the inclusion of weaker exons (de la Mata et al., 2003; Robson-Dixon and Garcia-Blanco, 2004). In addition, splicing can also have effect on the upstream events, such as chromatin modification, and mRNA export as well as translation (Kim et al., 2011). This crosstalk which exists between the splicing machinery and the continuum of gene expression raises the possibility that disruption of splicing process can have determinative effects on other regulatory processes within the cell, and thereby affects the phenotype of the human disease through indirect mechanism(s).

The mutations of the splicing machinery itself can give rise to pathological conditions such as spinal muscular atrophy (SMA) and autosomal dominant retinitis pigmentosum (adRP) (Padgett, 2012). SMA is caused as a result of mutation in *SMN2* splicing gene (Lefebvre et al., 1995) and sdRP is caused by the mutations of several spliceosomal genes (*PRPF31*, *PRPF8*, *HPRP3*, *PAP1* and *SNRNP200*) (Mordes et al., 2006). The *SMN2* gene product is necessary for maturation of the spliceosomes snRNPs and the mutations leading to a reduced level of this protein cause tissue specific death of spinal motor neurons (Lefebvre et al., 1995; Neuenkirchen et al., 2008). Recently mutations in the *U4atac* gene were identified as a cause of MOPD1 which is a multisystemic disease with growth retardation, abnormal bone morphology and severe brain malformation. U4atac snRNA is the RNA component of the minor spliceosome snRNP. Most of the *U4atac* mutations affect a highly conserved stem–loop structure required for association of the 15.5 K protein that mediates binding of proteins necessary for catalysis of trans-esterification reactions in the splicing pathway (Nottrott et al., 2002).

In addition, mutations within the consensus splice-site junction sequences between exons and introns of the target gene have also been reported, and are estimated to account for about 10% of human heritable disorders.

1.5.1 Splicing factor-3 b1 (*SF3B1*)

SF3B1, a major component of the U2 snRNP, is a highly conserved 155-kDa protein which along with other splicing factors including *SF3b2*, *SF3b3*, *SF3b4*, *SF3b5*, *SF3b14* and *PHF5A* forms a multiprotein complex called SF3b. *SF3B1* is the largest subunit of the *Sf3b* complex and plays an essential role in the recruitment as well as the stable binding of the U2 snRNP to the BPS near the 3' end of the intron, and helps to specify the site of splicing. In addition to being a component of the major spliceosome, *SF3B1* also plays a major role in the minor U12-dependent spliceosome.

SF3B1 is composed of two regions: the N-terminal domain (NTD, 1- 430AA) and the carboxy-terminus (431-1304AA). The N-terminal region of *SF3B1* contains several TPR repeats and, functions as a scaffold to facilitate interactions with other splicing associated factors such as *U2AF65*, *SF3b14*, *p14*, Cyclin-E and *Nipp-1* (Cass and Berglund, 2006; Yokoi et al., 2011). The *p14* protein which is highly conserved in a wide variety of organisms forms a stable complex with *SF3B1* and this protein complex acts as a core responsible for the BPS recognition. Importantly, *p14* and *SF3B1* complex interacts directly with the BPS adenosine and the flanking regions on the target mRNA, respectively (Gozani et al., 1998; Query et al., 1996; Will et al., 2001). In addition to this, *U2AF1* has also been suggested to interact with *SF3B1*; however the region and the mechanism of the interaction is not known (Gozani et al., 1998). The carboxy terminal (C-terminal) of *SF3B1* is highly conserved and contains 22 non-identical tandem repeats of the HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, target of rapamycin 1) motif. A single HEAT motif contains 37-47 amino acid residues forming two anti-parallel α-helices and two turns. These HEAT repeats contain conserved arginine and aspartic residues, and short flexible linkers between repeats (Groves and Barford, 1999; Xing et al., 2006). This part of the protein mediates protein–protein interactions with proteins that are involved in mRNA translation, cell signalling and nucleo–cytoplasmic transport. Using electron microscopy, it has been demonstrated that *SF3B1* HEAT repeats wrap around the entire U2 snRNP, marking out an S-shaped path around the overall shell-like form that stabilizes the U2 snRNP complex (Figure 4) (Golas et al., 2003).

The three-dimensional structure of U11/U12 di-snRNP which also contains *SF3B1* has recently been determined by electron microscopy. The electron microscopy of SF3b

complex revealed that only one side of this complex is able fit well with the U11/U12 di-snRNP while as the other half of the complex showed a severe density mismatch in the region where SF3b complex interacts with U11/U12 di-snRNP. This is due to differences in the diameters of the outer wall of SF3b complex and U11/U12 di-snRNP. This observation led to the idea that SF3b complex has to undergo a conformational change in its second shell half in order to fit completely into the density of the U11/U12 di-snRNP. Previous studies have demonstrated that the HEAT domains of nuclear transport factor importin-β can adopt different degrees of curvature, depending on the size of its binding partner such as importin-alpha, RanBP2. This possibility for changing the curvature of Sf3b1 HEAT domains is needed for the structural rearrangement of SF3b to fit it completely into the U11/U12 di-snRNP. Therefore, the potential opening of the two shell halves of the Sf3b complex may thus be assumed to be driven by, or to drive, a change in the curvature of the Sf3b1 HEAT repeat.

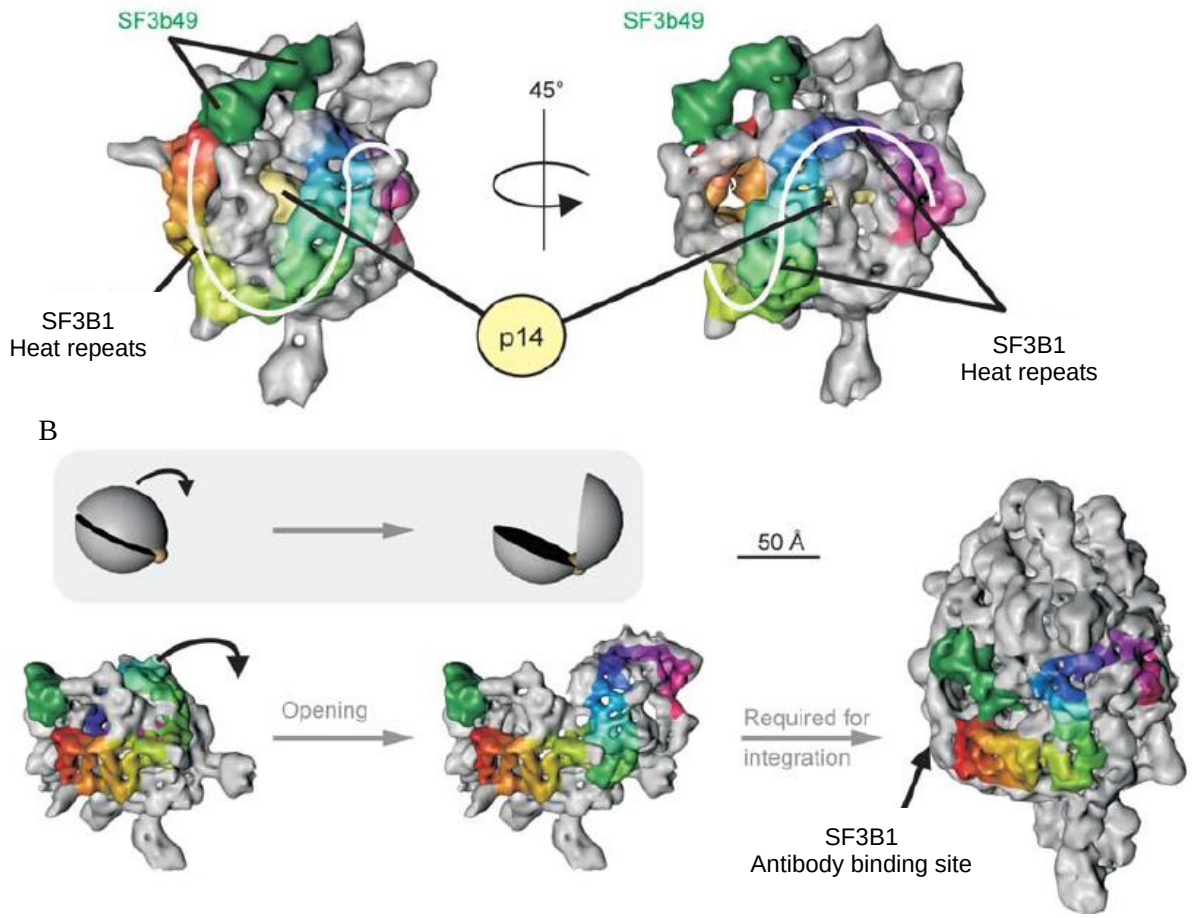


Figure 4: Three-dimensional structure of Sf3b complex and its structural rearrangement upon integration with U11/U12.

(A) Three out of seven proteins were located on the basis of their structural domains and their fit into the density map. The pattern of 22 tandem HEAT repeats of Sf3b1 were found wrapped around the SF3b complex in an S-shape. (B) In order for Sf3b to fit into the U11/U12 di-snRNP density, an opening of Sf3b is required. The Sf3b opening requires the movement of one shell-like half of the complex to match the outer wall of U11/U12 di-snRNP (Golas et al., 2005; Stark and Luhrmann, 2006).

Sf3b1 also contains several cdk consensus phosphorylation sites within its N-terminus. In *in-vitro* studies, it has been shown that Cyclin E-cdk2 uses Sf3b1 as a substrate and phosphorylation of Sf3b1 occurs prior or during the catalytic step II of the splicing pathway, suggesting a potential role for cdk regulation and a link to the cell cycle (Seghezzi et al., 1998). Furthermore, Sf3b1 also appears to be dephosphorylated by PP1/PP2A phosphatases concomitant with the second step of splicing. This indicates the importance of phosphorylation of Sf3b1 during splicing process which might act as a molecular switch that contributes to the fine tuning of the

splicing process and hence provide additional opportunities for regulation (Wahl et al., 2009).

1.6 Refractory anaemia with ring sideroblasts and Iron metabolism

The sideroblastic anaemia's are a heterogenous group of acquired and inherited disorders mainly characterized by the presence of ring sideroblasts in the bone marrow. The term "sideroblast" was first introduced by Kaplan to describe an erythoblast with a cytoplasm in which siderotic granules could be seen by ordinary microscope. However, in 1953 Douglas and Dacie described the presence of large siderotic granules in the form of a ring structure surrounding the nucleus of an erythoid precursor (Douglas and Dacie, 1953).

RARS, a subtype of MDS is an acquired form of sideroblastic anaemia, characterized by isolated anaemia, erythroid dysplasia and less than 5% bone marrow myeloblasts (Cazzola and Invernizzi, 2011). RARS is notable for more than 15% ringed sideroblasts which are basophilic and polychromatophilic erythroblasts characterized by accumulation of iron in the form of ferritin within the mitochondria that is visualized by Prussian staining as a blue perinuclear ring (Boultwood et al., 2008; Camaschella, 2009; Cazzola and Invernizzi, 2011). 5-year survival rates in RARS exceeds 50% and a risk of transformation to AML is low (Cazzola and Invernizzi, 2011; Malcovati et al., 2009; Novotna et al., 2008). RCMD-RS is phenotypically similar (i.e. ringed sideroblasts) to RARS but in addition 10% of non-erythroid lineage cells show dysplastic changes.

The main clinical features of RARS is severe anaemia (present in >90% of the patients), necessitating chronic blood transfusions, with a subsequent risk of iron overload. Some RARS patients have infections related to granulocytopenia or haemorrhage i.e. complications due to thrombocytopenia. Hepatomegaly and/or splenomegaly are rare and red cell volume is often increased. Peripheral blood film shows hypochromic or dimorphic red cells. Oval macrocytes, teardrop, fragments may also be present. Red cell elliptocytosis and basophilic stippling is also present to a variable degree.

Interestingly, majority of RARS patients have a normal karyotype. However, studies involving X-chromosome inactivation in female patients have shown that RARS arises

from hematopoietic stem cell with the potential for lymphoid and myeloid differentiation (Cazzola et al., 2000; Cazzola and Invernizzi, 2011; Malcovati et al., 2009).

Iron is an essential component for haem synthesis in maturing erythroblasts. It has the capacity to accept and donate electrons, therefore making it an effective component of oxygen-binding molecules, cytochromes, and a variety of other enzymes (Sheftel et al., 2009). However, this property can also lead to toxicity if its absorption, storage and transfer are not tightly regulated within the cell. Iron is transported within the serum bound to transferrin (Figure 5). Transferrin (Tf) binds to its receptor the transferrin receptor-1 (TfR1) on the cell surface. Receptor-ligand complex internalization via clathrin-coated pits ensues. Acidification of the endosome to pH 5.5 through the proton pump promotes the release of $Fe3^{+}$ from Tf that remains bound to TfR1 on the cell membrane (Ohgami et al., 2005; Ponka et al., 1998; Richardson et al., 2010). The ferrireductase Steap3 reduces $Fe3^{+}$ to $Fe2^{+}$ (Ohgami et al., 2005), which is transported by divalent metal ion transporter 1 (*DMT1*) from endosome to the cytosol or, possibly, directly to mitochondria in erythroid cells (Richardson et al., 2010). Iron is then exported to the mitochondria where it is incorporated into protoporphyrin IX to form haem, or incorporated into nascent iron–sulfur clusters (ISCs). The mechanism by which iron is exported to the mitochondria remains unclear. However, studies have suggested a ‘kiss and run’ hypothesis, according to which transferrin-containing endosomes in erythroid cells might be involved in the transfer of iron directly to the mitochondria (Richardson et al., 2010). The conserved cytosolic glutaredoxins Grx3 and Grx4 play an important role in intracellular iron sensing as well as trafficking, as their depletion in yeast leads to impaired iron transport to mitochondria and defects in iron-dependent pathways. SLC transporter mitoferrin (which is also known as *SLC25A37*) localizes to the inner mitochondrial membrane is also required for the entry of iron into mitochondria (Shaw et al., 2006). Alternatively, iron can be exported into the cytoplasm where it is incorporated into cytoplasmic ferroproteins or stored in cytosolic storage molecules, such as ferritin (Arosio et al., 2009). Free iron which is not bound to transferrin molecules usually accumulates within the cell when the iron-binding capacity of transferrin is exceeded. This free iron circulates as a complex mostly with citrate or acetate (Breuer et al., 2000).

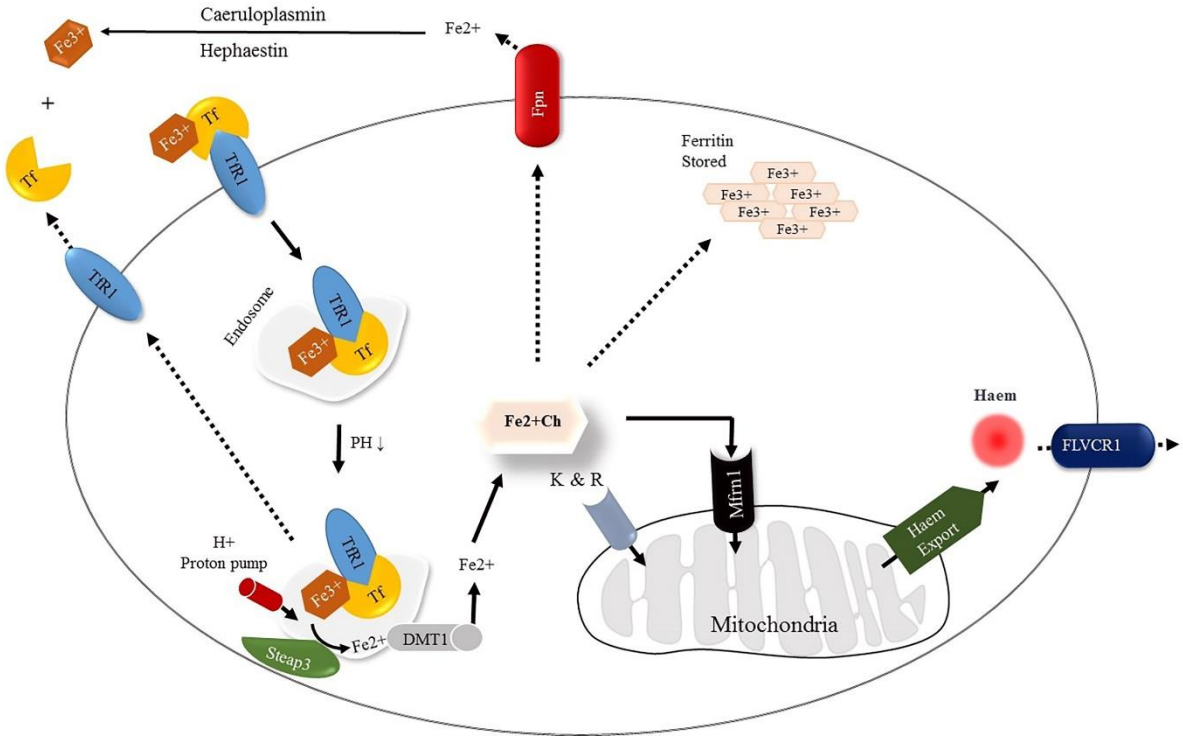


Figure 5: Iron traffic in erythroid progenitors that synthesise haem.

Fe^{3+} is transported across the cell membrane as diferric transferrin by the transferrin receptor (TfR1) and then taken by endosomes. A proton pump acidifies the endosomes, resulting in the release of Fe^{3+}. Steap3 reduces it to Fe^{2+}, which is then exported by DMT1 to the cytoplasm. Fe^{2+} exported by *DMT1* is delivered to mitochondrial mitoferrin-1 (Mfrn1) either by direct contact (the kiss-and-run mechanism, K&R) or through intermediate transport by as-yet uncharacterized cytoplasmic chaperones (Fe2þCh). Mitoferrin-1 imports iron into mitochondria where iron is incorporated into haem. Then, haem is exported via an unknown exporter (Haem export) and incorporated into globin chains to generate haemoglobin. Excess iron is stored in ferritin. Under some circumstances, iron is also exported as ferrous iron via ferroportin (Fpn) or as haem via feline leukaemia virus C receptor (FLVCR1). The apo-Tf–TfR1 complex is transported back to the cell surface, where apo-Tf is released to capture plasma Fe^{3+}.

The production of haem starts with a series of reactions catalysed by enzymes that produce protoporphyrin IX (Figure 6). In the mitochondrial matrix, the first step in protoporphyrin IX production involves the condensation of glycine and succinyl coenzyme-A resulting in the formation of 5-aminolevulinic acid (ALA). This reaction is catalysed by ALA synthase (ALAS) which resides within the mitochondrial matrix. The

universal precursor ALA is exported to the cytosol and converted to coproporphyrinogen III in a series of four enzyme-catalysed reactions, which is then transported back to the mitochondria. Coproporphyrinogen III is converted to protoporphyrinogen IX and then to PPIX within the mitochondria. The final reaction in the haem synthesis pathway is the addition of one atom of Fe 2^{+} into protoporphyrin IX and this step is catalysed by ferrochelatase (FC) located at the inner mitochondrial membrane. Haem is then exported to the cytosol for incorporation into haemoproteins through ABC family of transporters and the SLC transporter SLC25A39 (Napier et al., 2005).

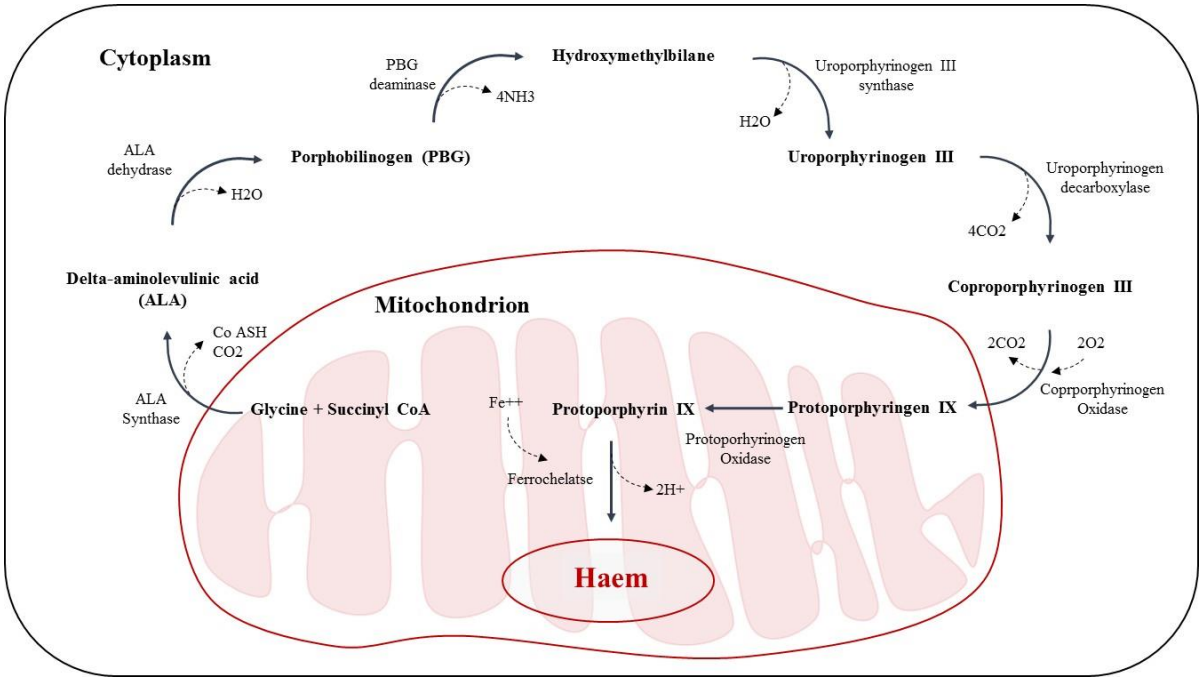


Figure 6: Haem biosynthesis pathway.

Haem synthesis is initiated by the synthesis of ALA by the condensation of the amino acid glycine and succinyl CoA (from kerbs cycle) by the pyridoxal phosphate-requiring enzyme namely δ-aminolevulinic acid (ALA) synthase. This is a rate-limiting reaction and it is negatively regulated by glucose and haem concentrations. Following synthesis, ALA dehydratase dimerizes two molecules of ALA in the cytoplasm to produce the pyrrole ring compound porphobilinogen. Hydroxymethylbilane synthase catalyses the head-to-tail condensation of four molecules of porphobilinogen to produce the linear tetrapyrrole intermediate called hydroxymethylbilane. Once produced, hydroxymethylbilane is converted to uroporphyrinogen III with the help of enzyme uroporphyrinogen-III synthase. Subsequently, uroporphyrinogen III is decarboxylated by the enzyme uroporphyrinogen decarboxylase and producing coproporphyrinogen III, a product with the methyl groups in place of acetate. Following its synthesis, coproporphyrinogen III is transported to the mitochondrion, where 2 propionate residues are decarboxylated with the help of oxygen-dependent enzyme coproporphyrinogen oxidase. This reaction produces protoporphyrinogen IX. In the mitochondrion, protoporphyrinogen IX is converted to protoporphyrin IX by protoporphyrinogen IX oxidase. The oxidase reaction requires molecular oxygen and results in the loss of 6 protons and 6 electrons, producing a completely conjugated ring system, which is responsible for the characteristic red colour of the haem molecules. The final reaction of the pathway also takes place in the mitochondrion and involves the insertion of the ferrous iron (Fe^{2+}) atom into the ring system generating haem with the help of enzyme ferrochelatase. Co-ASH, Coenzyme A.

In addition to the haem synthesis, iron is also required for the biogenesis of all nuclear iron-sulphur protein clusters (ISCs) within the mitochondria (Figure 7). ISCs are metallocofactor prosthetic groups of multiple proteins involved in various metabolic reactions and are critical for the normal functioning of both mitochondrial proteins (respiratory chain complexes, ferrochelatase, aconitase) and cytosolic proteins (xanthine oxidase). ISCs are formed with the help of the mitochondrial ISC assembly machinery, which combines iron and cysteine (source of sulphur) to produce ISCs on scaffold proteins, and finally incorporates these clusters into the recipient apoproteins (Apo) (Lill and Muhlenhoff, 2008). Initial assembly of ISCs on the Isu1/2 (scaffold proteins) requires that the two elemental components of the machinery are present in their reduced state (Zheng and Dean, 1994). These components include the iron-binding protein Frataxin (putative iron donor) and the cysteine desulfurase complex Nfs1 in complex with *ISD11* (Lill and Muhlenhoff, 2008). Nfs1 converts cysteine to alanine and persulfidic sulphur which in-turn is reduced to sulphur, presumably through the electron transport chain comprised of ferredoxin reductase and ferredoxin (Fdx), which receive electrons from NADH. The release, transfer and incorporation of ISCs into the recipient apoproteins is mediated by the ISC assembly machinery components such as Hsp70 chaperone Ssq1 and the monothiol glutaredoxin *GLRX5*; however, their mode of action remains unclear (Lill and Muhlenhoff, 2008; Sheftel et al., 2009; Wingert et al., 2005). Impaired Fe-S cluster biogenesis in *GLRX5* RNAi cells and *GLRX5*-deficient patient cells leads to mitochondrial iron overload and cytosolic iron depletion (Ye et al., 2010).

The Fe-S protein biogenesis also occurs outside the mitochondria in the cytoplasm and requires the components of the mitochondrial ISC export machinery and the core ISC assembly machinery. To date, various components involved in this process have been identified, such as the P-loop NTPase Cfd1 (cytosolic Fe–S cluster-deficient protein 1), Dre2, Nar1 (nuclear architecture-related protein 1)/IOP1 (iron-only hydrogenase-like protein 1), Cia1 and Tah18 (Netz et al., 2007). The ATP binding cassette transporter protein *ABCB7* also acts as a central component in the biogenesis of ISCs.

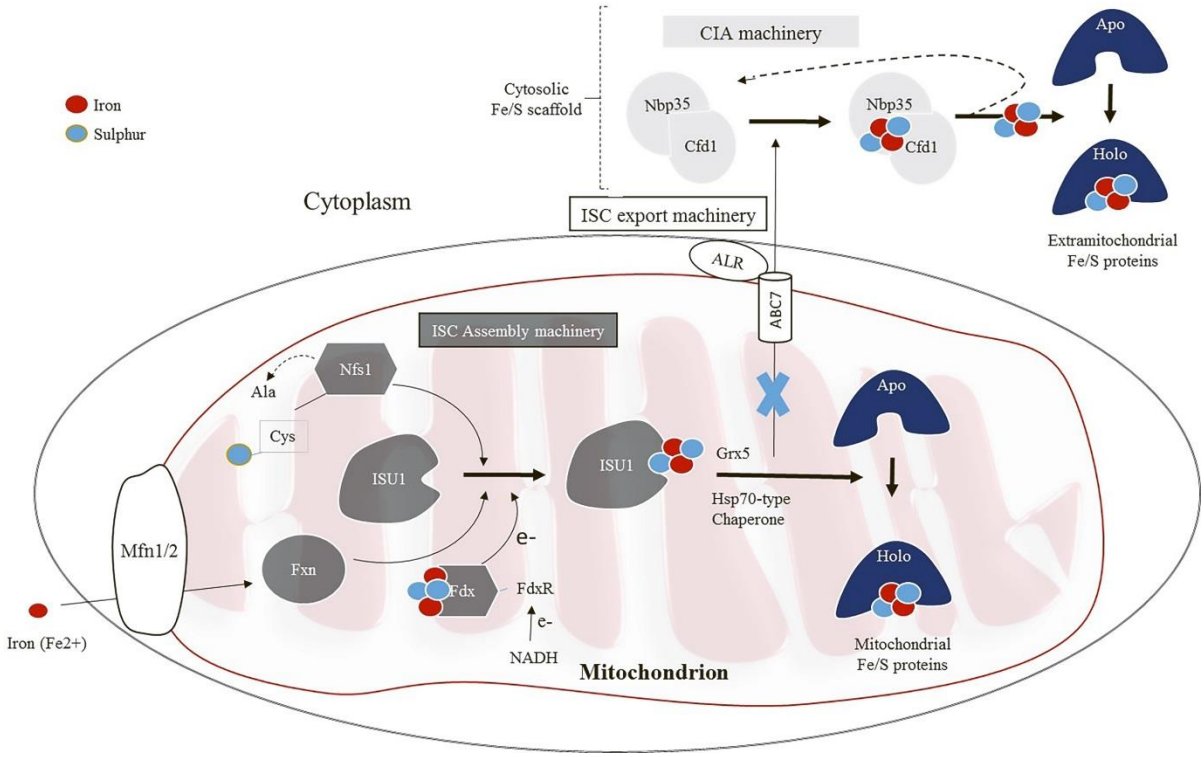


Figure 7: Schematics of Fe-S protein biogenesis.

Fe-S protein biogenesis involves the crucial function of mitochondria which import iron in ferrous (Fe2+) form from the cytoplasm in a membrane potential dependent manner. This Import is facilitated by the inner membrane carriers Mfn1/2. Maturation of mitochondrial Fe-S holoproteins (Holo) involves two major steps. The synthesis of a transiently bound Fe-S cluster on the scaffold protein Isu1 which occurs with the help of the components of the mitochondrial iron-sulfur cluster (ISC) assembly machinery. These proteins include the cysteine desulfurase complex Nfs1, Fdx-FdxR complex which serves as the sulfur donor for cluster synthesis, and receive their electrons (e−) from NADH. The release of the Fe-S cluster from Isu1 and its transfer as well as incorporation into recipient apoproteins (Apo) are facilitated by late components of the ISC assembly machinery (dark arrows). Fe-S cluster release from ISU11 is achieved by interaction of the Hsp70 chaperone. The monothiol glutaredoxin Grx5 is involved in this second major step. The extramitochondrial Fe/S protein biogenesis requires the core ISC assembly machinery and components of the mitochondrial ISC export machinery. The ABC7 transporter of the inner membrane along with ALR exports an unknown compound (X, Blue) to the cytoplasm for use in Fe-S protein assembly. In the cytoplasm, the cytoplasmic Fe-S protein assembly (CIA) machinery catalyzes Fe/S protein maturation. First, Fe/S clusters are assembled on the P-loop NTPases Cfd1 and Nbp35 complex. The Fe-S clusters are bound to Cfd1-Nbp35 in a labile fashion and subsequently incorporated with into cytosolic and nuclear apoproteins. Fxn = Frataxin; X (blue) = unknown substrate exported by ABCB7. Cfd1/Nbp35 are a cytosolic scaffold complex.

The expression of key proteins involved in the iron metabolism such as *ALAS2*, Ferritin, and TfR1, is regulated by intracellular iron. TfR1 and ferritin are also regulated post-transcriptionally through the binding of iron responsive protein 1 (IRP1) or iron responsive protein 2 (IRP2) to iron responsive elements (IREs), therefore maintaining cellular iron homoeostasis (Figure 8) (Anderson et al., 2012). IREs are evolutionarily conserved hairpin structures of 25–30 nucleotides present within the UTRs of their respective mRNAs. The assembly of Fe-S cluster on IRP1 reduces its RNA binding affinity while increasing its aconitase activity, therefore providing one of the mechanisms for regulating intracellular iron levels. In addition, defects in any of the three Fe-S protein biogenesis systems; ISC assembly machinery, ISC export machinery, and the cytosolic Fe-S protein assembly (CIA) machinery also effects the assembly of Fe-S cluster on IRP1 and alters IRE binding activity. Previous studies have shown that the translation of the transcripts is blocked when IRPs bind to IREs motifs in 5´-UTR of the respective mRNA (such as ALAS2, DMT1 and HIF-2*α*). However, the stabilization as well as translation of mRNA transcripts (such as MRCK*α* and Cdc14A) proceeds when IRPs bind to IRE motifs in 3´-UTR of the target mRNAs (Wang and Pantopoulos, 2011).

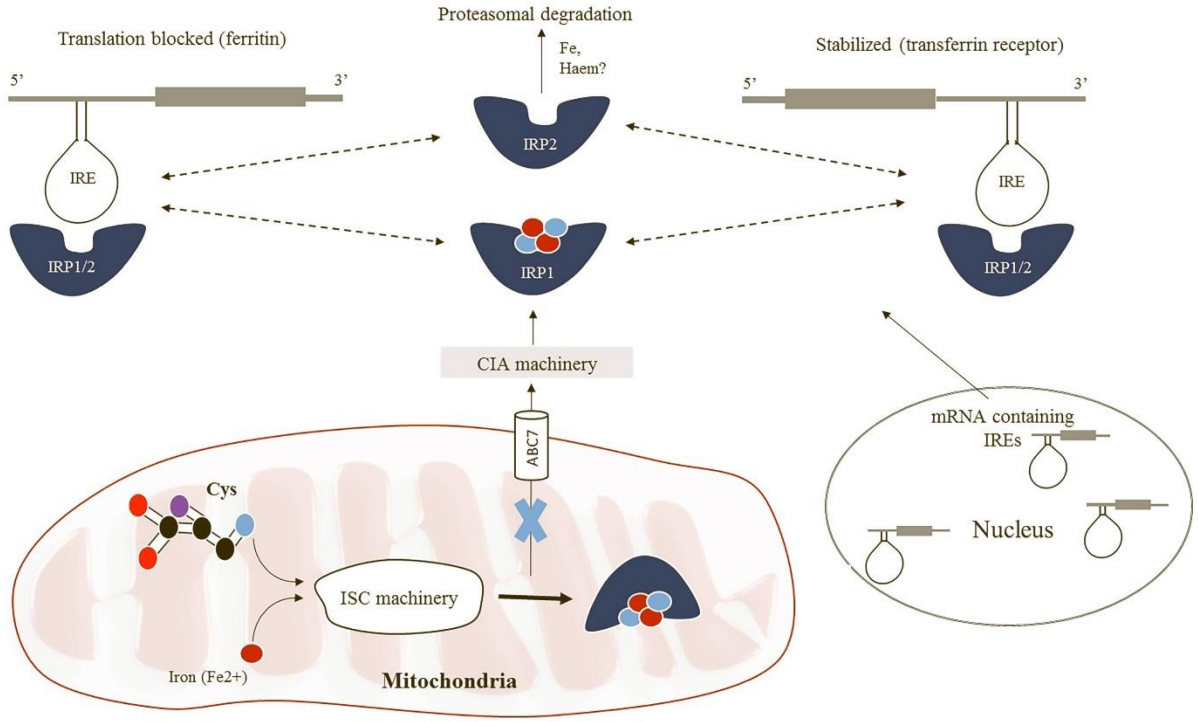


Figure 8: Schematics of Fe-S cluster on IRP1 and IREs.

The presence or absence of a Fe-S cluster on IRP1 serves as a binary switch, determining the binding capacity of the protein to IREs. IRP1 binds to IREs in the absence of Fe-S cluster, thereby stabilizing mRNAs with 3'-UTR IREs and blocking translation of mRNAs with 5'-UTR IREs. Therefore, appropriate regulation of iron handling requires the mitochondrial ISC machineries and also a functional cytosolic assembly system (CIA). X (blue) = unknown substrate exported by *ABCB7*, IRE- Iron regulatory elements, IPR- Iron regulatory protein.

Ring sideroblasts are also characteristic of other inherited sideroblastic anaemias such as X-linked sideroblastic anaemia (XLSA), sideroblastic anaemia with spinocerebellar ataxia (XLSA/A) and Pearson's syndrome (Cazzola and Invernizzi, 2011). Abnormalities of several genes involved in haem biosynthesis, iron-sulfur (Fe-S) cluster biogenesis, Fe-S cluster transport, and mitochondrial metabolism have been implicated in these disorders. Mutation of *ALAS2* (Bergmann et al., 2010; Cotter et al., 1999), *ABCB7* (Maguire et al., 2001), and also mitochondrial DNA deletions (pearson syndrome) (Smith et al., 1995) are all associated with bone marrow ring sideroblast. Interestingly, none of these candidate genes have been found to be mutated in RARS. Increasing evidence suggests mitochondrial dysfunction occurs in early stages of erythroid maturation that can lead to abnormal mitochondrial iron deposition and

ineffective haemopoiesis. Gene expression profiles of $CD34^+$ cells from RARS patients shows significant upregulation of mitochondria-related genes and in particular, genes involved to haem synthesis pathway (such as *ALAS2*) (Pellagatti et al., 2006). Defective *ALAS2* enzyme activity in bone marrow erythroid cells results in insufficient synthesis of protoporphyrin IX, mitochondrial iron overload, and intramedullary death of red cell precursors. Moreover, mitochondrial ferritin gene encoded by *FTMT* has also been reported to be highly over-expressed in patients with acquired sideroblastic anaemia (Nikpour et al., 2010). In contrast, *Boultwood et al* in another study reported substantial down-regulation of iron/sulphur cluster transporter *ABCB7* gene in patient with RARS (acquired) (Boultwood et al., 2008). *ABCB7* is required for normal hematopoiesis and siRNA knockdown of the gene in HeLa cells results in iron-deficient phenotype with mitochondrial iron overload.

Using *in-vitro* erythroid differentiation assay, *Tehranchi et al* showed that the early erythroid progenitor cells from low-risk MDS patients (including RARS patients) spontaneously release excessive amounts of cytochrome C from mitochondria, resulting in activation of caspase-9 and subsequent cell death (Tehranchi et al., 2003). Moreover, enhanced sensitivity to Fas ligation was abrogated by inhibiting caspase-9 activity. Therefore, increased sensitivity of MDS progenitors to death receptor stimulation appears to be due to a constitutive activation of the apoptotic signalling pathway in the malignant cells. A recent study has reported that Fas-caspase-8 pathway and Fas-associated protein with death domain (FADD) are involved in erythroid apoptosis (Claessens et al., 2005). In contrast to inherited sideroblastic anaemia where ring sideroblasts are found only in the late erythroblasts, in acquired RARS erythroblasts accumulate ferritin within the mitochondria during early differentiation, before morphological signs of erythroid differentiation are visible (Tehranchi et al., 2005). These findings support the notion that acquired sideroblastic anaemia's are unrelated to a defect in haem biosynthesis directly, however, it might be due to an iron metabolism alteration that manifests itself in early erythroblasts. Thus, the genetic abnormalities present in RARS patients must both give a proliferative advantage leading to expansion of the malignant clone as well as contribute to the ineffective erythropoiesis and abnormal erythroid iron metabolism, leading to accumulation of the iron in the mitochondria of erythroid progenitors. It is

likely that this defect involves a gene(s) or a pathway which is responsible for the regulation of various cellular process such as erythroid differentiation, mitochondrial function, or cellular iron metabolism.

Abnormal epigenetic profile, such as aberrant DNA methylation and histone modification, may also contribute to the altered gene expression and genomic instability, suggesting that disease occurs as a result of multiple and cumulative alterations to the hematopoietic stem cells. However, whether chromosomal aberration, gene mutations or epigenetic changes represent the initiation or secondary events, remains unclear.

1.7 Therapeutic strategies for RARS

Immunosuppressive regiment's and thalidomide, which may have proven effective in subsets of MDS patients such as RA, are ineffective in patients with RARS (Park et al., 2008). Lenalidomide has been demonstrated to abrogate transfusion requirement in approximately 25% of the patients with EPO-refractory RARS (median duration of response 43 weeks) (Jadersten et al., 2008). The ineffective haemopoiesis in RARS responds less well to EPO than other low risk subgroups of MDS such as RA (7.5% vs. 21%). However, the combination of EPO and granulocyte colony-stimulating factor (G-CSF) induces durable responses (overall response rate of 50%) (Ingram et al., 2006). Various *in-vitro* and *in-vivo* studies have shown that G-CSF strongly inhibits apoptosis in differentiating RARS erythroblasts by inhibiting the release of mitochondrial cytochrome-C, with a subsequent decrease in caspase-9 and caspase-3 activity (Schmidt-Mende et al., 2001; Tehranchi et al., 2003; Tehranchi et al., 2005). However, G-CSF had no effect on the accumulation of mitochondrial ferritin at any time point of stem cell differentiation (Tehranchi et al., 2005). Moreover, EPO has been previously demonstrated to inhibit apoptosis (Koury and Bondurant, 1990). Interestingly, patients with RARS have significantly better survival and higher response rate (survival time - 121+ months, response rate 71%) to EPO/G-CSF as compared to RCMD-RS patients (survival time - 31 months, response rate 30%) (Howe et al., 2004; Jadersten et al., 2008). This difference correlates well with the predictive model for erythroid response, since RCMD-RS patients usually belong to intermediate IPSS category, whereas RARS patients belong to a good risk group (Hellstrom-Lindberg and Cazzola, 2008).

1.8 High throughput molecular technologies

High-throughput molecular technologies such as second-generation sequencing (SGS) and microarrays have accelerated the pace for the identification of novel genetic mutations/variations and transcripts associated with various cancers including haematological malignancies.

1.8.1 Second Generation Sequencing

Since the completion of the human genome project, major changes have occurred in the approach to genome sequencing that has moved away from traditional Sanger sequencing (first generation sequencing). SGS has allowed highly streamlined and multiplex sample preparation prior to DNA sequencing of individual clones in a massively parallel fashion. This provides significant time savings and a minimal requirement for associated equipment in comparison to automated, multistep pipelines necessary for clone-based traditional high-throughput sequencing. The basis of SGS is the amplification of hundreds to millions of single DNA molecules separately but in the same experiment (in parallel). The DNA to be sequenced is used to prepare a library of fragments that have oligonucleotide adapters (synthetic DNAs) attached to each end of the DNA fragment by use of DNA ligase. These universal adapters are specific to each sequencing platform that can be used to polymerase-amplify the library fragments during specific steps of the protocol. These library fragments are then amplified in situ on a solid surface, either a bead or a flat glass microfluidic channel with attached adapter sequences that are complementary to the universal adaptors on the library fragments. The SGS amplification process allows each amplified fragment to yield a single focussed molecule (cluster of amplified DNA fragments that originated from a single DNA molecule). SGS is often referred to as massively parallel due to its large scale and high throughput of samples, which is an appropriate term to describe the technology that follows fragment amplification to yield large sequencing data in a limited amount of time. This highly efficient massively parallel sequencing technology involves a stepwise reactions that consist of (1) a nucleotide addition to the DNA fragment, (2) a detection step to determine the identity of the incorporated nucleotides, and (3) a wash step that includes removal of fluorescent labels or blocking groups attached to the incorporated nucleotide. Therefore, SGS instruments performs sequencing and detection simultaneously which

allows thousands to billions of reaction foci to be sequenced at the same time. However, the yield of sequence reads from SGS instruments varies from several hundred thousand reads (Roche/454, Illumina MiSeq) to tens of millions of reads (Illumina HiSeq2000 and Applied Biosystems SOLiD), thus bringing close to the time when sequencing an individual's genome quickly and cheaply is feasible.

However, there are particular challenges for the detection and diagnosis of cancer genome abnormalities. For example, somatic mutation calling is more complex than germline mutation calling because some acquired alterations in cancer genome are prevalent at a low frequency (<20% of cells) in clinical samples, often owing to substantial admixture with normal cells. In addition, a key parameter defined for each mutation is its allelic burden (or mutant allele burden) which is defined as the expected fraction of reads in the tumour cells that harbour the mutation among all reads that map to the same genomic location. Due to this admixture problems, it is difficult to determine the accurate mutant allele burden (MAB) of the acquired genomic alternation through the traditional Sanger sequencing. However, SGS technology can solve such problems and capture the local complexity of the tumour genome as well as the non-tumour contamination levels with high accuracy.

Sequencing of complete AML genome using SGS technology represented a landmark in genetic screening and has since led to the identification of significant disease markers such as *IDH1/2* and *DNMT3A* mutations. However, SGS coupled to new DNA-enrichment technologies, such as array/bead hybridization-based region capture, has enabled highly efficient targeted sequencing of regions of interest, such as disease-linked genes, in greatly increasing numbers of samples simultaneously. Applying SGS technology to perform 'whole-exome' sequencing (WES) allows an unbiased investigation of the complete protein-coding fraction of the human genome. Recently, comprehensive exome 'capture arrays' have become available for selectively fractionating approximately 50 Mb or ≈ 200,000 coding exons across the human genome. A clear limitation of WES is that it does not identify the noncoding variants found by whole-genome sequencing. However, efficiency benefits of selectively sequencing the exome in isolation means at least 20 times as many samples can be processed at one time compared to whole-genome sequencing, Furthermore, data arising from WES is directly attributable to cellular function and thus

'bioinformatically' less challenging. Various studies involving WES have demonstrated that this approach is a cost-effective means of precisely denoting functional variations in a large cohort of patient samples. Recent studies involving exome-sequencing through capillary-based sequencing in colorectal, breast and pancreatic carcinomas, and glioblastoma, led to the discovery of somatic mutations in *IDH1* in glioblastoma and germline mutations in the gene encoding partner and localizer of *BRCA2* (*PALB2*) in patients with pancreatic carcinoma. Studies have also successfully applied targeted capture/SGS to identify genetic changes involved in Mendelian diseases such as Miller syndrome. Furthermore, in a study of Nephronophthisis-related ciliopathies (NPHP-RC), 12 different truncating mutations of *SDCCAG8* (serologically defined colon cancer antigen 8) were identified through WES in 10 families affected by NPHP-RC, therefore demonstrating the benefits of targeting sequencing in such diseases.

1.8.2 Microarray analysis

Over the past decade, microarray technology has also led the way from studying the individual biological function of a gene, protein or, at best, pathways towards more global analysis of cellular activity. In addition, this technology has improved our understanding of the underlying mechanisms of the malignant process involved in various cancers.

The publication of whole-genome sequencing studies raised an important question about the molecular interpretation of the somatic/inherited gene mutation, and not so much the number of mutated genes, that are responsible for the variation between various cancers. In addition, alternative splicing is another well-established post-transcriptional modification that plays an essential role in regulating gene expression and has been reported to have a role in the pathogenesis of haematological malignancies (Oltean and Bates, 2014).

Gene expression profiling (GEP) provides a detailed image of the transcriptional activity of whole genome at a specific time point. Golub and colleagues were the first to demonstrate a difference between AML from acute lymphoblastic leukaemia (ALL) patients based on their global gene expression (Golub et al., 1999). Furthermore, GEP in diffuse large B-cell lymphoma (DLBCL) led to the classification of patients according to their expression profiles and as a result DLBCL were divided into two subtypes: activated B-cell-like DLBCL and germinal center B-cell-like DLBCL. These two

subtypes differ in the proliferation rate, differentiation level, response to therapy and prognosis (Alizadeh et al., 2000). In addition, genome wide gene-expression profiling has provided an insight into the molecular pathogenesis of breast cancer which has enabled to predict disease recurrence and response to treatments as well as gain insights into oncogenic pathways and the process of metastatic progression (Liu et al., 2007; van 't Veer et al., 2002).

With recent developments in microarray technology, splicing sensitive microarrays that contain probes targeted to individual exons or exon-junctions are becoming increasingly popular. By targeting individual exons, exon junctions, and annotated isoform variants, such microarrays enable us to profile not only the expression levels of the entire transcript, but also detect the transcriptional representation of exons (i.e. splice variant). These microarray chips contains a set of unique and distinct 25-mer oligonucleotides probes that are immobilised to a solid surface or membrane on the chip. These probes are designed such that each probe binds to specific sequence corresponding to an exon, splice variant or a gene. The microarray probes are labelled with a detectable fluorochrome molecule. The level of binding between a specific probe and its target sequence is quantified by measuring the fluorescence when scanned by the instrument. This signal, in turn, provides a measure of the expression of the specific gene or splice variant containing the target sequence. The latest generations of microarray chips interrogate 285,000 transcripts as well as analyse individual exons of genes separately (GeneChip® Human Transcriptome Arrays 2.0), or even more advanced, the entire genome for transcriptional activity without focusing on predicted genes (Tiling arrays).

One of the most frequently utilized methods in studying of the gene expression as well as splicing is Affymetrix GeneChip® technology (Santa Clara, CA, USA). Affymetrix' GeneChip® Human Transcriptome high-resolution array contains >6.0 million distinct probes covering coding and non-coding transcripts with 70% of the probes on this array cover exons for coding transcripts, and the remaining 30% of probes on the array cover exon-exon splice junctions and non-coding transcripts.

1.9 Aims

Over the last 15 years, there have been significant advances in understanding the molecular pathogenesis of MDS. However, establishing the relevance of individual gene aberrations found in MDS with a clinical phenotype, particularly RARS, has been a major challenge. Applying SGS technology to perform whole-exome sequencing on in low-risk MDS with relatively homogenous phenotype such as MDS patients with ring sideroblasts will allow a thorough and unbiased investigation of the complete protein-coding regions. This should enable the unravelling of the genetic make-up and elucidate the pathophysiology, refine the prognostic scoring systems, and more importantly provide novel therapeutic targets for this subgroup of MDS. This study will therefore provide a valuable insight into the potential molecular mechanisms associated with this sub group of MDS patients.

- The main aim is to identify disease specific mutation(s)/variation(s) in a phenotypically well-defined MDS subgroup RARS/RCMD-RS with ring sideroblasts using DNA-capture (whole-exome) and SGS technology.
- Validate a relevant candidate(s) variation/mutation profile in a larger cohort of MDS patients to understand the clinical/statistical significance within the disease group.
- To functionally characterize the selected functionally relevant candidates in *in-vitro/ in-vivo* studies using appropriate cell lines and/or mouse model.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Patient Samples

Twelve patients (MDS patients with >15% ringed sideroblasts, n =11; Congenital sideroblastic anaemia, n=1) were initially selected for whole-exome sequencing (Table 6). 156 MDS patients (Table 7) were selected for follow-up mutation analysis of the splicing pathway genes, epigenetic modifiers and other cell signalling/transcription regulator genes, followed by validation of the candidate mutations. Patients with MDS seen at King's College Hospital from June 2004 to June 2013 were enrolled in this study. All patients had provided written informed consent in accordance to King's College London Haemato-Oncology Tissue Bank research ethics protocol (08/H0906/94). Demographic and clinical characteristics of the studied patients are detailed in Table 7. All patients were risk stratified according to IPSS categories. The clinical variables, FAB, WHO subtype and the prognostic risk of all patients were ascertained at the time of sample collection. The median follow-up was 21.4 months (range, 1-83 months). The patient cohort was followed up to January 2012 for disease progression, and survival. The survival data for patients who underwent allogeneic haematopoietic stem cell transplant (HSCT) (n=35, 22%) were censored on the day of transplant.

Table 6: Clinical characteristics of 11 MDS-RS patients and one congenital sideroblastic anaemia patient used for whole-exome sequencing.

BM blasts- Represent bone marrow blasts at the time of sample; M- Male; F- Female; N/A- Not available; Int- Intermediate.

Patient ID	Sample Tim Point	Sample Time Point	Age	Gender	WHO Diagnosis	Treatment Prior To Sample	BM Blasts (%)	Haemoglobin	Lymphocytes	Platelet Counts	Ringed Sideroblast %	IPSS	Cytogenetics	Transfusion Dependant	Survival Status	AML Evolution	Disease Progression
MDS 1	TP1	1	77	M	RARS-T	Epo/GCSF	2	8.8	9.53	268	>15	Low	NORMAL	Yes	A	No	No
MDS 1	TP2	2	-		RARS-T	Epo/Len	2	10.5	1.9	248	>15	Low	NORMAL	Yes	-	No	No
MDS 2	TP1	1	75	M	RCMD-RS	Epo/GCSF	1	8.6	1.87	166	41	Low	NORMAL	Yes	-	No	Yes
MDS 2	TP2	2	-	-	RCMD-RS	Epo/Len	1	7.1	3.67	137	50	-	NORMAL	Yes	-	No	Yes
MDS 2	TP3	3	-	-	sAML	Epo/Len	90	7.5	1.63	29	-	-	46,XY,add(7)(q32),add(10)(p13)[7]/46,XY,add(7)(q22),add(7)(q32),add(10)(p13)[3]	Yes	D	Yes	Yes
MDS 3	N/A	N/A	63	M	RARS	Epo/GCSF	1	9.1	3.15	570	>15	Low	NORMAL	No	A	No	No
MDS 4	N/A	N/A	58	F	RARS	None	1	10.7	3.21	196	24	Low	NORMAL	No	A	No	No
MDS 5	N/A	N/A	22	M	Congenital sideroblastic anaemia	None	1	10.6	4.7	400	32	N/A	NORMAL	Yes	A	No	No
MDS 6	N/A	N/A	74	M	RCMD-RS	Epo/Len/ Aza	1	12.1	3.44	151	35	Low	NORMAL	No	A	No	No
MDS 7	N/A	N/A	67	M	RARS	ATG	0	10.4	4.5	358	50	Low	NORMAL	Yes	A	No	No
MDS 8	N/A	N/A	58	M	RARS	None	0	7.7	0.9	281	50	Int-I	NORMAL	Yes	A	No	No
MDS 9	N/A	N/A	60	M	RARS	Epo/GCSF	1	9.8	5.3	334	47	Low	NORMAL	No	A	No	No
MDS 10	N/A	N/A	63	F	RARS-T	None	1	10.5	6	565	80	Low	NORMAL	No	A	No	No
MDS 11	N/A	N/A	73	M	RARS	Epo/GCSF/ ATO	0	9.7	2.80	236	78	Low	NORMAL	Yes	A	No	No
MDS 12	N/A	N/A	67	F	RARS	EPO/GCSF	2	7.1	1.7	389	>35	Int-I	NORMAL	Yes	A	No	No

Table 7: Clinical characteristics and treatments received by all 156 MDS patients studied for targeted amplicon sequencing.

Cytogenetics failed in 4 patients in this cohort. † denotes other treatments which include; lenalidomide, thalidomide, cyclosporine and Antithymocyte globulin (ATG). * includes 13 patients who had HSCT after receiving 5-azacitidine. n- Represents number of patients. % - Represents percentage of the patients.

Patient Characteristics	Overall
Total patients	156
Age, years	
Median	65.5
Range	17-85
Sex	
Male [n (%)]	105 (67%)
Female [n (%)]	51 (33%)
WHO category *	
RA/RCMD [n (%)]	40 (26%)
RARS/RCMD-RS [n (%)]	26 (17%)
RAEB -1/2 [n (%)]	49 (31%)
sAML [n (%)]	15 (10%)
tMDS/AML [n (%)]	12 (8%)
CMML &MPD/MDS-U [n (%)]	14 (9%)
Bone marrow blasts	
Median (%)	5
Range	0-80
IPSS cytogenetic risk group	151
Good [n]	92
Intermediate [n]	17
Poor [n]	44
Transfusion dependency	
Yes [n (%)]	80 (51%)
Progression to AML	
Yes [n (%)]	45 (29%)
Treatments	
5-Azacytidine	68
Other Treatments†	14
HSCT*	35
No treatment	39

2.1.2 Commercial Kits

2.1.2.1 Agilent SureSelect Target Enrichment kit for Illumina paired-end sequencing library (v. 2.0.1)

Agilent's SureSelect Target Enrichment System is a highly efficient hybrid selection technique used for enriching targeted regions of the genome from repetitive sequences and non-exonic sequences. It encompasses all exons annotated by the GENCODE project as well as Consensus CDS and RefSeq databases. This kit has the following components; blocking reagents, hybridization buffers, wash and elution buffers, nucleotides, SureSelect biotinylated RNA library "BAITS" and other reagents.

2.1.2.2 TruSeq PE Cluster Kit v3-cBot-HS for Illumina paired-end sequencing

The TruSeq PE Cluster Kit v3-cBot-HS provides reagents for the cBot cluster amplification system for illumina paired-end sequencing. The cluster generation is performed on cBot system (Illumina). This cluster generation kit has the following components; primers, blocking reagents, DNA polymerase and other buffers.

2.1.2.3 TruSeq SBS Kit v3-HS for Illumina paired-end sequencing

TruSeq SBS v3-HS kits contain ready-to-load reagents for accurately determining the DNA sequence on Illumina HiSeq2000 platform. This kit has the following components; wash buffer, incorporation mix buffer, high salt buffer, incorporation wash buffer, cleavage buffer, cleavage mix reagent, enhanced scanning mix reagent, enhanced DNA polymerase (EDP), long read nucleotide mix (FFN) and NAOH.

2.1.2.4 Genome Sequencer (GS) FLX Titanium Series Kits

GS FLX Titanium series kit contains all reagents required for emulsion-based clonal amplifications of DNA libraries and for subsequent sequencing on a Genome sequencer FLX instrument. This kit has the following components: LV emPCR Kit; emPCR breaking Kit LV/MV 12pc; picoTiterPlate Kit 70x75; sequencing kit XLR70; control bead kit.

2.1.3 Buffers

Phosphate Buffered Saline (PBS)

PBS tablet (Oxoid) 1 tablet; Sterile distilled water 100ml

Phosphate Buffered Saline with Tween-20 (PBST)

Sterile PBS 1L; Tween 20 (VWR) 0.02% (v/v)

1X Tris-Borate-EDTA (TBE)

800ml of H2O; 200ml of 20X TBE

1X Tris-EDTA (TE)

990ml of H2O; 10ml 1MTris-Hcl; 400µl 0.25M EDTA

Magnetic activated cell sorting (MACS) buffer

Phosphate-buffered saline (PBS) 500ml (pH 7.2); Bovine serum albumin (BSA) 0.5% (w/v); EDTA 2mM

Fluorescent Activated Cell Sorting (FACS) buffer

Sterile PBS (pH 7.2) 500ml; Fetal calf serum (Sigma) 2.5% (v/v); Sodium Azide (VWR) 0.1% (v/v)

Freezing Medium

Foetal calf serum 45ml; Dimethyl Sulfoxide (DMSO, Sigma) 4.5ml

70% Ethanol

dH2O 15mls; Ethanol 35mls

80% Ethanol

dH2O 10mls; Ethanol 40mls

1M NaOH (Melt solution)

NaOH (10M) 0.125 ml; dH2O 9.875 ml

2.2 Methods

2.2.1 Extraction of Genomic DNA

DNA was extracted from $CD34^+$ cells (n=12), $CD34^-CD3^+$ (n=18), $CD34^-CD235^+CD71^+$ (n=3), $CD34^-CD235^-$ (n=3), $CD34^-CD3^+CD4^+$ (n=3), $CD34^-CD19^+$ (n=3), skin (n= 27), buccal swab (n=3) and bone marrow total nucleated cells (n=156) using QIAamp DNA extraction kit (Qiagen, UK) according to the manufacturer's protocol. Cells were pelleted by centrifugation at 2000xg for 5 minutes and then the supernatant was discarded. Cells were resuspended in 200µl of phosphate buffered saline *(*PBS) and then transferred into the 1.5ml tube. 20µl of Qiagen protease and 200µl of buffer AL was added to the same 1.5ml tube and mixed by pulse vortexing for 15 seconds. The sample was incubated for 1 hour at 56^{o}C. 200µl of 100% (v/v) ethanol was added to the sample tube. The sample was mixed on a votex for 15 seconds and then transferred to the Qiagen spin column. The spin column was centrifuged for 1 minute at 8000xg. Then the Qiagen spin column was transferred into clean collection tube and 500µl of buffer AW1 was added to it, followed by centrifugation for 1 minute at 8000xg. Qiagen spin column was transferred to a clean collection tube and 500µl of buffer AW2 was added followed by centrifugation for 3 minutes at 13000xg. The spin column was then transferred to a clean 1.5ml tube and 100µl of elution buffer was added followed by centrifugation at 8000xg for 1 minute. DNA was stored at +4^{o}C.

Out of 156 cases, whole genome amplified (WGA) DNA was used in 40 cases for targeted amplicon sequencing.

2.2.2 DNA Quantitation using Quant-iT picoGreen dsDNA Assay kit.

DNA concentration was determined by using the Rotor-Gene 6000 and Quant-iT pico-green dsDNA Assay kit. Quant-iT™ PicoGreen® dsDNA reagent (Life Science Technologies) is an ultra-sensitive fluorescent nucleic acid stain for quantitating double-stranded DNA (dsDNA) in solution. 1µl of DNA was diluted in 10µl of 1x Tris-EDTA (TE) buffer (Life Science Technologies) and 10µl of 1x florescent dye pico-green (Life Science Technologies) was added to the tube containing the DNA. Then the tubes containing the sample mixture were run on the Corbett Rotor-gene RG300A alongside with the 6 dsDNA standards (Life Science Technologies) with

concentrations of 2ng/µl; 1ng/µl 0.5ng/µl; 0.25ng/µl and 0.125ng/µl respectively. The data was analysed by using the Rotor-gene software.

2.2.3 RNA Extraction

Upto 1 x 10^7 cells ($CD34^+$ and $CD71^+CD235^+$) were used for RNA extraction using RNeasy RNA extraction mini kit (Qiagen, UK) according to the manufacturer's protocol. Cells were lysed using 1ml of Trizol (Thermo Fisher Scientific, UK) and incubated for 5 minutes at room temperature. 0.2 ml of chloroform was added to the lyzed cells and vortexed for 15 seconds. Cells were centrifuged for 10 minutes at 15000xg (at 4^{o}C). 200µl of the supernatant was transferred into the fresh 1.5ml eppendoff tube. 700µl of Qiagen RLT buffer was added to the mixture. 500µl of 100% ethanol was added to the mixture. Samples were vortexed for 15 seconds. Samples were transferred to the Qiagen MiniElute spin columns. Spin column tubes were centrifuged for 15 seconds at 10000xg. Spin columns were transferred into a new collection tubes. 500µl of RPE was added to the spin column and centrifuged for 15 seconds at 10000xg. Then, 750µl 70% ethanol (v/v) was added to the spin columns and centrifuged for 15 seconds at 10000xg. Spin columns were centrifuged again for 5 minute at 15000xg. The spin column was then transferred to a clean 1.5ml tube and 10µl of RNAse free H_2O was added to the spin columns followed by centrifugation at 8000xg for 1 minute. RNA was stored at -20^{o}C.

2.2.4 Whole-exome Sequencing

For the generation of standard exome capture libraries, Agilent SureSelect Target Enrichment reagent (Agilent Technologies, UK) and protocol (Figure 9) for Illumina paired-end sequencing library was used.

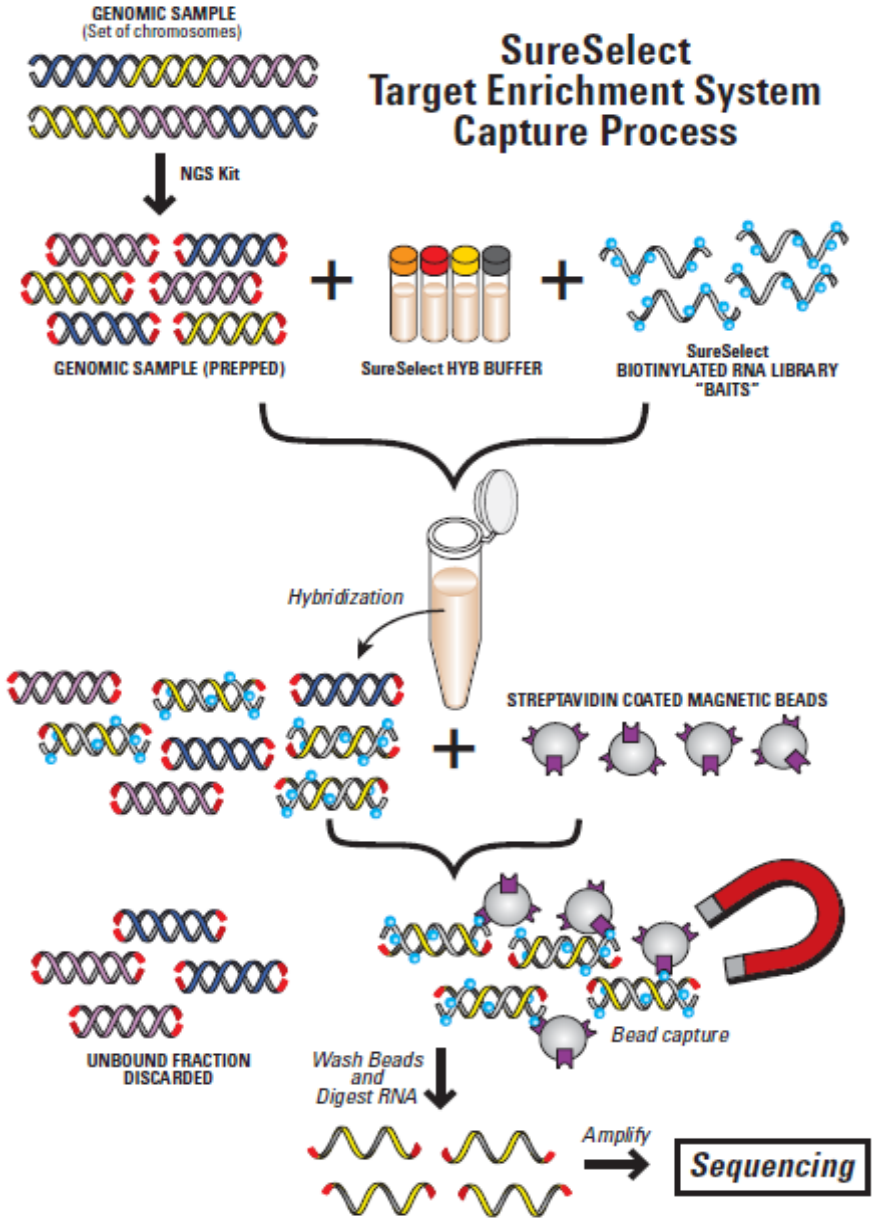


Figure 9: Schematic representation of the work-flow for whole-exome capture.

Following the sonication/shearing of the patient genomic DNA, Illumina adapters are ligated on the DNA fragments. Agilent's RNA baits and streptavidin coated magnetic beads are used to capture the exonic regions of the DNA. Unbound fraction of the genome which doesn't bind to the RNA baits is washed away using various buffer washes. Illumina sequencing adapters are used to amplify the target exome. Prepared exonic library is then sequenced on Illumina HiSeq2000.

2.2.4.1 DNA Sonication

DNA sonication was performed using Bioruptor (Diagenode, Belgium) or Covaris ultrasonicator E220 (Covaris, Inc. USA). Prior to sonication, patient DNA samples were dissolved in a sonication buffer 1x TE with a varying DNA concentration (80ng to 1000ng). DNA sonication was performed according to the optimized conditions as shown in Table 8 and Table 9.

Table 8: Bioruptor sonication protocol for gDNA.

Tissue type	Target size	Bath Temperature	Time	Cycle Condition (On/Off cycle time)	Cycle number	Power	Sample Volume
$CD34^+$	200 to 300 bp	4ºC - 8ºC	12	30 Sec/30 Sec	4	Low	100 µl
Skin Biopsy, $CD3^+$	200 to 300 bp	4ºC - 8ºC	13.25	30 Sec/30 Sec	4/5	Low	100 µl

Table 9: Covaris ultrasonicator sonication protocol.
Sonication protocol for genomic DNA (gDNA) and Whole genome amplified (WGA) DNA.

Settings	Value
DNA Target Peak	150-200 bp
Duty Factor	10%
Peak Incident Power (PIP)	175
Cycles per Burst	200
Treatment time	360 Seconds
Bath Temperature	4ºC - 8ºC
Sample Volume	120 µl

2.2.4.2 AMPure Bead purification

Fragmented DNA was purified using Agencourt AMpure XP beads (Beckman Coulter*)* by following the manufacturer's protocol. AMpure XP magnetic beads were added to fragmented DNA in 1:1.8 ratio (Beads:DNA) and the mixture was incubated at room temperature for 20 minutes. Beads were washed twice by using the magnetic particle concentrator (MPC) and 500µl of 70% (v/v) ethanol. DNA was eluted by resuspending the beads in 35µl of nuclease free H_2O.

2.2.4.3 Pre-Hybridization library preparation

The purified DNA fragmented library for each patient was end-repaired by adding end-repair enzyme (Agilent Technologies, UK) cocktail (Appendix Table 1) to the fragmented DNA library. Samples were incubated in a thermal cycler for 30 minutes at 20^{o}C without heated lid. Following end-repair, samples were purified using Agencourt AMpure XP beads as described in section 2.2.4.2. DNA libraries were eluted in 32µl of nuclease free H_2O.

End-repaired DNA fragments were tailed with poly dA's by adding a Klenow exo^{-} enzyme (Agilent Technologies, UK) cocktail (Appendix Table 2) to the samples and incubated for 30 minutes at 37^{o}C with a heated lid at 50^{o}C. The libraries were purified using Agencourt AMpure XP beads as described in section 2.2.4.2. DNA samples were eluted in 29µl of nuclease free H_2O. Illumina adapters were ligated to the fragmented DNA library using the InPE adapter oligo mix (Illumina), DNA ligase (Agilent Technologies, UK) and 5x T4 DNA ligase buffer (Appendix Table 3; Agilent Technologies, UK). Samples were incubated in a thermal cycler for 1 hour at 20^{o}C without heated lid. Ligated DNA fragmented libraries were purified using Agencourt AMpure XP beads as described previously in section 2.2.4.2. The adaptor-ligated libraries were eluted in 37µl of nuclease free H_2O.

Adapter ligated DNA libraries were amplified using Herculase II fusion DNA polymerase enzyme (Agilent Technologies, UK) cocktail (Appendix Table 4). Polymerase chain reaction (PCR) amplification was performed as mentioned in appendix table 5. Amplified libraries were purified using Agencourt AMpure XP beads as described in section 2.2.4.2. DNA samples were eluted in 32µl of nuclease free H_2O.

2.2.4.4 Hybridization/Exome Capture

The exomic regions of the genome were enriched using Agilent SureSelect Human All Exon Kit (Agilent Technologies, UK) covering approximately 50Mb of the genome and Agilent post capture buffer kit (Agilent Technologies, UK). This target enrichment process involves using ultra-long 120-mer biotynilated cRNA baits which are used to capture the region of interest i.e. exomic regions from the genomic fragmented library. 500ng of amplified adaptor ligated library from each patient was mixed with hybridization buffers (Appendix Table 6), blocking mixes (Appendix Table 7) and 7µl of diluted SureSelect all exon capture library (5µl of SureSelect capture library mixed with 2µl of diluted RNase block) in a step wise manner, according to the Agilent SureSelect Target Enrichment protocol (Agilent Technologies, UK). Samples were incubated in a thermocycler for 24 hours at 65°C with heated lid.

50µl Dynal MyOne streptavidin T1 (Invitrogen) magnetic beads were used for each exome capture library. Beads were washed 3 times by using MPC and 200µl SureSelect binding buffer (Agilent Technologies, UK). Following on, Dynal MyOne streptavidin T1 beads were resuspended in 200µl of SureSelect binding buffer.

After 24 hours hybridization of the adaptor-ligated DNA libraries, samples were directly transferred from the thermal cycler to the tubes containing the washed Dynal MyOne streptavidin T1 beads. Hybrid-capture/bead solution was incubated for 30 minutes at room temperature. After incubation, hybrid-capture/beads were washed by using MPC and 500µl of SureSelect wash buffer 1 (Agilent Technologies, UK). Hybrid-capture/beads were then washed 3 times by adding 500µl of pre-warmed SureSelect wash buffer 2 (Agilent Technologies, UK), incubating for 10 minutes at 65°C, followed by MPC separation. Targeted DNA from the exomic regions was recovered by resuspending hybrid-capture/beads in 50µl of SureSelect elution buffer (Agilent Technologies, UK). Supernatant was transferred into a fresh 1.5ml tubes and 50µl of SureSelect neutralization buffer (Agilent Technologies, UK) was added to the captured library. Captured DNA libraries were then purified by using Agencourt AMpure XP beads as described in section 2.2.4.2. DNA libraries were eluted in 30µl. The captured exomic DNA fragments were amplified through PCR by using Herculase 11 fusion DNA polymerase (Agilent technologies, UK) cocktail (Appendix Table 8). Index tags/primers were used to barcode each patient in order for them to be sequenced

together. PCR amplification was performed as mentioned in appendix table 9. Following amplification, samples were purified using Agencourt AMpure XP beads as described in section 2.2.4.2. DNA libraries were eluted in 30µl of nuclease free H_2O. Exomic DNA index-tagged library concentration was determined by 2 separate quantification assays inorder to obtain the precise concentration for efficient data generation. Initially exomic DNA index-tagged library concentration was determined by using the Rotor-Gene 6000 and Quant-iT pico-green dsDNA Assay kit as described in section 2.2.2. Later, exomic DNA index-tagged library concentration was determined by using Stratagene brilliant-111 Ultra-Fast SYBR Green qPCR assay (Agilent Technologies, UK). Based on the concentration values obtained from Quant-iT pico-green dsDNA Assay, exomic DNA index-tagged libraries were diluted in duplicate using 1x dilution buffer. The qPCR consisted of 10µl of 2X Brilliant master mix, 0.4µl of Illumina primer mix, 0.3µl of reference dye, 2µl of diluted DNA index-tagged library and 7.3µl of dH_2O. PCR was performed on 7900HT Fast Real Time PCR system using the manufactures conditions (Appendix Table 10). Following on, exomic DNA index-tagged library concentrations were calculated by using 7900HT Fast Real Time PCR software. Exonic DNA index-tagged libraries were pooled together in equal concentrations depending upon the indexes used for the patient samples.

2.2.4.5 Illumina HiSeq2000 Sequencing

Pooled exomic DNA index-tagged libraries were diluted to 2nM using 1x TE with 0.1% (v/v) tween-20. Pooled exomic libraries were denatured by adding 10µl of 0.1N (w/v) NAOH to 10µl of pooled exomic libraries and incubated at room temperature for 8 minutes. Following on, 980µl of pre chilled HT1 buffer (Illumina) was added to the 20µl of denatured template pooled exomic libraries. Denatured libraries were further diluted to 11pM using HT1 buffer (Illumina) and then mixed on a vortex. 120µl of denatured libraries were transferred into an eight-tube strip and placed on ice.

Cluster generation of the prepared libraries was performed using a TruSeq PE Paired-End Cluster Generation kit v3 (Illumina) and cBOT cluster generation system (Illumina). Cluster generation process involves the hybridization of millions of DNA templates to a lawn of oligonucleotides immobilized on the flow cell surface.

Immobilized DNA templates are amplified by isothermal bridge amplification. Cluster generation reagent plate (Illumina), HiSeq flow cell (Illimina) and tubes containing the denatured DNA libraries were loaded on the cBOT cluster generation system. Clustered HiSeq flow cell was removed from the cBOT cluster generation system upon completion of the cluster generation and stored overnight in the 50ml tube containing the storage buffer (Illumina) at 4ºC.

Paired end sequencing with a read length of 2x 75bp was performed using TruSeq SBS reagents v3 kit (Illumina) and Illumina HiSeq2000 system. The Illumina HiSeq2000 uses sequencing-by-synthesis technology and produce more than 1 billion quality reads or around >350 gigabases (100bp paired-end reads) of data per run. TruSeq SBS reagents tray and clustered HiSeq flow cell was loaded on the Illumina HiSeq2000 instrument. After performing the required pre-run checks on the Illumina HiSeq2000, sequencing read 1 was started using the Illumina HiSeq2000 control software.

After the read 1 sequencing was complete, paired end reagents (HiSeq read 2 cluster re-synthesis kit) for read 2 were loaded on the Illumina HiSeq2000. Following on, sequencing read 2 was started using the Illumina HiSeq2000 control software.

2.2.4.6 Data Generation, Alignment Pipeline, SNP detection and Structural variation calling

Whole-exome sequencing data was processed by using an in-house data analysis pipeline (Figure 10). Illumina RNASeq 2000 Real Time Analysis (RTA) or the Off-Line Basecaller (OLB) software generates the *.bcl files during the sequencing run. *.bcl files were de-multiplexed and converted to fastq using CASAVA1.8.2 (Illumina). Sequence alignment to the human reference genome (NCBI37/Hg19) and variant calling for SNPs and Indels was performed using both CASAVA 1.8 (Illumina) and Burrows-Wheelers aligner (BWA) (Li and Durbin, 2009) as well as GATK (McKenna et al., 2010) pipelines according to the Broad Institute best practices pipeline. Sequence alignment was performed by Dr. Alexander Smith.

For the illumina pipeline, Annovar software (Wang et al., 2010b) was used to perform functional annotation of all variants. Following on, variants with a quality score of

>QSNP 25 and with a tumour read depth of >2 read (total read depth > 6) were passed for further analysis.

For BWA/GATK pipeline, VarScan 1.3.4 (Koboldt et al., 2009; Koboldt et al., 2012) was used to call somatic variants on pileup files (Samtools) (Li et al., 2009) for paired tumour and normal (skin or $CD3^+$ T-cells) tissue. Reads with alignment score of <10, base quality score <15 and fold strand bias >10 were excluded. Pindel (Ye et al., 2009) was also used to screen for more computationally challenging large Indel mutations at specific loci such as for FLT3 ITD. All resulting variants were passed through ANNOVAR (Wang et al., 2010b) using refseq annotation. Genomic duplicated regions were also filtered out unless associated with known mutations. Somatic mutations were subsequently passed when somatic p-value (VarScan, Fisher's exact T-test) was < 0.01, had ≥3 reads supporting mutation, were present in paired normal tissue at less than 20% of the paired tumour tissue and had a allele burden of >5%.

Candidate variants from both pipelines obtained were passed for validation if the variants were deemed to cause protein changes and not found in dbSNP140, esp5400 and 6500, and 1000 genomes databases at >0.01 population allele frequency.

A repeat sequencing run was also used to filter out sequencing artefacts where candidate mutations supported in both runs were passed. Additionally, variations present in available paired skin and $CD3^+$ T-cell samples at >20% (skin) or >50% (T-cells) the level found in $CD34^+$ tissue were excluded from subsequent analysis.

Data from both pipelines were largely in agreement for filter passed variants and variant depth >2. Any discrepancies between the two pipelines at this level were later confirmed to be artefactual calls. Any known myeloid specific variant which didn't pass our filtration criteria was included in further analysis.

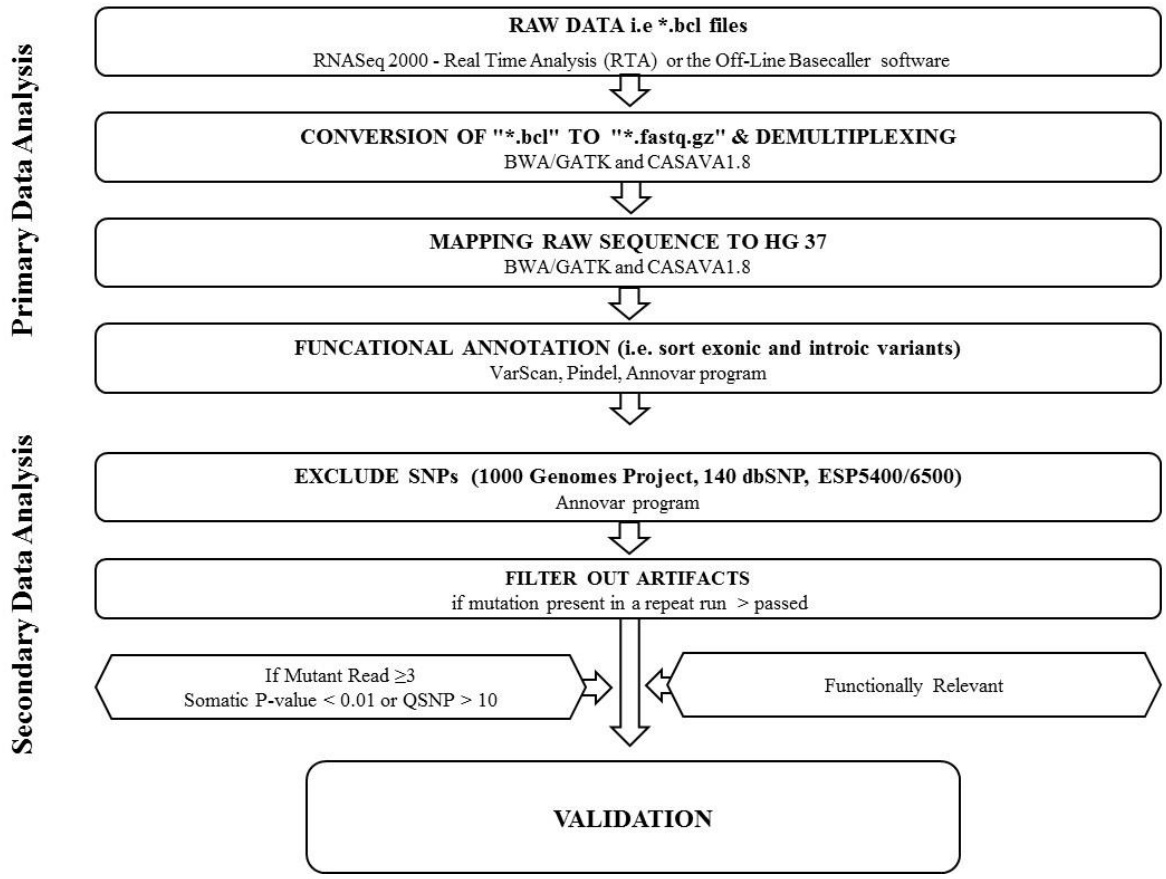


Figure 10: WES data processing pipeline for detecting somatic mutations.

2.2.4.7 Data filtration criteria for Xenograft samples

For isolated human engrafted cells (HEC) $CD45^{+}CD33^{+}$ from mice, candidate mutations were only passed if they were present in more than 1 HEC sample or found concurrently in one or more HEC experiment and long-term tissue culture (LTC) derived sample or primary $CD34^{+}$ cells for the same patient. Where only WGA-derived data from HEC was available for one engraftment experiment, or a combination of WGA and non-WGA data from more than one engraftment experiment, mutations were only passed when they were found concurrently in either: (1) WGA HEC and non-WGA HEC independent experiments; (2) Repeat WGA DNA samples from the same engraftment experiments or; (3) WGA HEC experiments and LTC derived sample, $CD34^{+}$ or HEC experiments derived from non-WGA DNA. For WGA-derived HEC, a stand-alone analysis was also performed for individual experiments. For this, additional false positive filters were applied to remove additional WGA-derived

artefacts with positional and read length distribution bias as well as variant and mapping quality biases. Variants with homopolymer runs of five or more flanking the mutation were also filtered out. WGA-derived variants from control WGA experiments were also removed. For MDS3, only variants found in at least two of the three animal experiments from WGA-derived data were included in final list due to increased artefactual events. Stand-alone analysis of non-WGA HEC cells were likewise performed but only using filters up to step 7 above.

2.2.4.8 Mapping Loss of Heterozygozity (LOH) Across Chromosomes by WES

LOH was mapped from WES data according to the following criteria: Variants were called as described above. Variants found in 1000 genome project database with a population allele frequency less than 0.001 were selected and filtered out for known genomic duplicated regions, read depth <20 and somatic status according to VarScan. The filtered set of variants was then plotted graphically for the required chromosome.

2.2.4.9 Exome Coverage

Exome coverage was calculated utilizing GATK Depth Of Coverage script against intervals that the Agilent V4 Sureselect kit capture probes were based on. For non-WGA sample, mean base coverage across the targeted region ranged from 77-115X and for >98% of the exome base coverage was >10 reads in all cases, based on 9-13 gigabase of sequence data. For the same amount of sequence data on average, WGA-derived exome data yielded average base coverage of 55-75X and 65-83% (average 74%) of exome was covered ≥10 reads.

2.2.5 454 Roche Amplicon Sequencing

2.2.5.1 Gene panel library preparation 3.2.4.1

Mutation screening for *SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*, *TP53*, *FLT3*, *DNMT3A*, *ASXL1*, *EZH2*, *NRAS*, *KRAS*, *JAK2*, *CCBL*, *RUNX1*, *CEBPA*, *BRAF*, *MPL*, *NPM1*, *IDH1*, *IDH2*, C-*KIT* and *TET2* was performed on bone marrow total nucleated cell DNA using the Roche GS FLX platform. All coding exons were selected for sequencing for genes including *ZRSR2*, *TP53*, *DNMT3A*, *EZH2*, *RUNX1*, *CEBPA* and

TET2. Mutation data of *TET2* for 142 cases that was available from our previously published study (Smith et al., 2010) was included in the analysis. Mutational hotspots were specifically sequenced for *SF3B1*, *SRSF2*, *U2AF1*, *ASXL1*, *CCBL*, *FLT3*, *NRAS*, *KRAS*, *JAK2* exon12/14, *IDH1*, *IDH2*, *BRAF*, *MPL*, *C-KIT* and *NPM1* in all cases. Gene hotspot regions were selected based on the previously published data and frequency of the mutations shown in COSMIC database (Table 10). In addition, all coding exons of *SF3B1* were sequenced for all 24 RARS/RCMD-RS patients present within our patient cohort. *SRSF2*, *U2AF1* and *ZRSR2* genes were amplified by using PCR primers published previously (Yoshida et al., 2011b). Specific PCR primers were designed to amplify and sequence coding regions or mutation hotspot regions of 22 genes (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*, *TP53*, *FLT3*, *DNMT3A, ASXL1, EZH2, NRAS, KRAS, JAK2, CCBL, RUNX1*, *CEBPA*, *BRAF*, *MPL*, NPM1, *IDH1, IDH2, C-KIT* and *TET2*). Primer3 program was used to design the PCR primers. The primers were tagged with a universal sequencing primer (aUS1 for forward; aUS2 for reverse). PCR primers for all the genes are shown in appendix table 11. First round PCR reaction was performed by using master mixes from Table 11 and using thermal cycler protocol as described in Table 12 (see table legends for more details). The amplicons were amplified using either Promega GoTaq Hot Start Green (Promega, Madison, WI) or Qiagen HotStarTaq DNA Polymerase (Table 12). The product size and specificity for each amplicon (for 5 patients) was confirmed by gel electrophoresis.

Table 10: Genes screened for hotspot mutations.
Mutation frequency within the hotspot regions was obtained from Cosmic database (Welcome Trust Sanger Institute, http://www.sanger.ac.uk /genetics/CGP/cosmic/).

Gene	Total Frequency of Mutations in MDS	Hot Spot Regions Sequenced	COSMIC Database
SF3B1	30%	Exon 12-16	Shows 99% of SF3B1 somatic mutations in haematopoietic neoplasms within this region (344 mutant /2668 sequenced samples)
SRSF2	15%	Exon 1	Shows 100% of SRSF2 somatic mutations in haematopoietic neoplasms within this region (279 mutant /1872 sequenced samples)
ASXL1	15%	Exon 12	Shows 99% of ASXL1 somatic mutations in haematopoietic neoplasms within this region (472 mutant / 3590 sequenced samples)
U2AF1	12%	Exon 2, 6	Shows 99% of U2AF35 somatic mutations in haematopoietic neoplasms within this region (96 mutant /1518 sequenced samples)
FLT3	6%	Exon14, 15	Shows 100% of FLT3 somatic mutations in haematopoietic neoplasms within this region (11816 mutant /51551 sequenced samples)
NRAS	3.6%	Exon 2, 3	Shows 99% of NRAS somatic mutations in haematopoietic neoplasms within this region (676 mutant /6473 sequenced samples)
JAK2	3%	Exon12, 14	Shows 100 % of JAK2 somatic mutations in haematopoietic neoplasms within this region (32218 mutant /76213 sequenced samples)
CCBL	2.3%	Exon 7, 8, 9	Shows 99% of CCBL somatic mutations in haematopoietic neoplasms within this region (163 mutant /3577 sequenced samples)
IDH2	2.1%	Exon 4	Shows 100 % of IDH2 somatic mutations in haematopoietic neoplasms within this region (905 mutant /13777 sequenced samples)
NPM1	1.8%	Exon 12	Shows ≥ 97 % of NPM1 somatic mutations in haematopoietic neoplasms within this region (4181 mutant /14167 sequenced samples)
IDH1	1.4%	Exon 4	Shows 100 % of IDH1 somatic mutations in haematopoietic neoplasms within this region (778 mutant /16301 sequenced samples)
KRAS	0.9%	Exon 2, 3	Shows >97% of KRAS somatic mutations in haematopoietic neoplasms within this region (152 mutant /3716 sequenced samples)
BRAF	0.5%	Exon 11, 15	Shows 100 % of BRAF somatic mutations in haematopoietic neoplasms within this region (6 mutant /581 sequenced samples)
MPL	Rare	Exon 10	Shows ≥ 97 % of MPL somatic mutations in haematopoietic neoplasms within this region (534 mutant /14202 sequenced samples)
C-KIT	Rare	Exon 17	Shows 40 % of C-KIT somatic mutations in haematopoietic neoplasms within this region (1672 mutant /7734 sequenced samples)

Table 11: PCR reagents and concentrations used to make master-mix A and master-mix B.

Master-mix A was used to amplify *ZRSR2, TP53, DNMT3A, EZH2, TET2, SF3B1, U2AF1, ASXL1*, *CBL, FLT3, NRAS, KRAS, JAK2* exon12/14, *IDH1*, *IDH2*, *BRAF, MPL*, *KIT* and *NPM1*. Master-mix B was used to amplify *SRSF2, RUNX1* and *CEBPA*.

Reagent	Volume per reaction (µl)
MASTER MIX A	
Nuclease Free H_2O	3
Hot start Go-taq master mix	5
DNA (@5ng/µl)	3
Primer (@ 5 uM)	1
Total volume	10
MASTER MIX B	
Nuclease Free H_2O	2.55
Buffer	1
$MgCl_2$	1
DMSO	0.75
dNTP Mix	0.5
Taq	0.2
Primer (@ 5 uM)	1
DNA(@5ng/µl)	3
Total volume	10

Table 12: Thermal cycler conditions used for the PCR.
Program A was used for *ZRSR2, TP53, DNMT3A, EZH2 , TET2, SF3B1, U2AF1, ASXL1, CBL, FLT3, NRAS, KRAS, JAK2* exon12/14, *IDH1*, *IDH2, BRAF, MPL, KIT* and *NPM1*. Program B was used *SRSF2, RUNX1* and *CEBPA*.

Steps	Temperature	Time
Program A		
Step 1 - Denaturation	95 °C	5 Minutes
Step 2 - Denaturation	95 °C	35 Seconds
Step 3 - Annealing	60 °C	40 Seconds
Step 4 - Extension	72 °C	45 Seconds
	Repeat Step 2 through step 4 for a total of 35 cycles	
Step 5 - Final Extension	72 °C	15 Minutes
Step 6 - Hold	4 °C	hold
Program B		
Step 1 - Denaturation	95°C	2 Minutes
Step 2 - Denaturation	95°C	35 Seconds
Step 3 - Annealing	60°C	40 Seconds
Step 4 - Extension	68°C	45 Seconds
	Repeat Step 2 through step 4 for a total of 35cycles	
Step 5 - Final Extension	68°C	15 Minutes
Step 6 - Hold	4 °C	hold

Second round PCR was then performed on 1 to 1.5µl of the amplified product obtained from the first round PCR (Table 11 & Appendix Table 12). This second round PCR is used to incorporate different 20 to 30bp multiple-identifiers (MIDs) within the DNA template for the identification of each patient (Figure 11). Patient-specific “barcodes” (MIDs) were derived from Roche GS Titanium protocol.

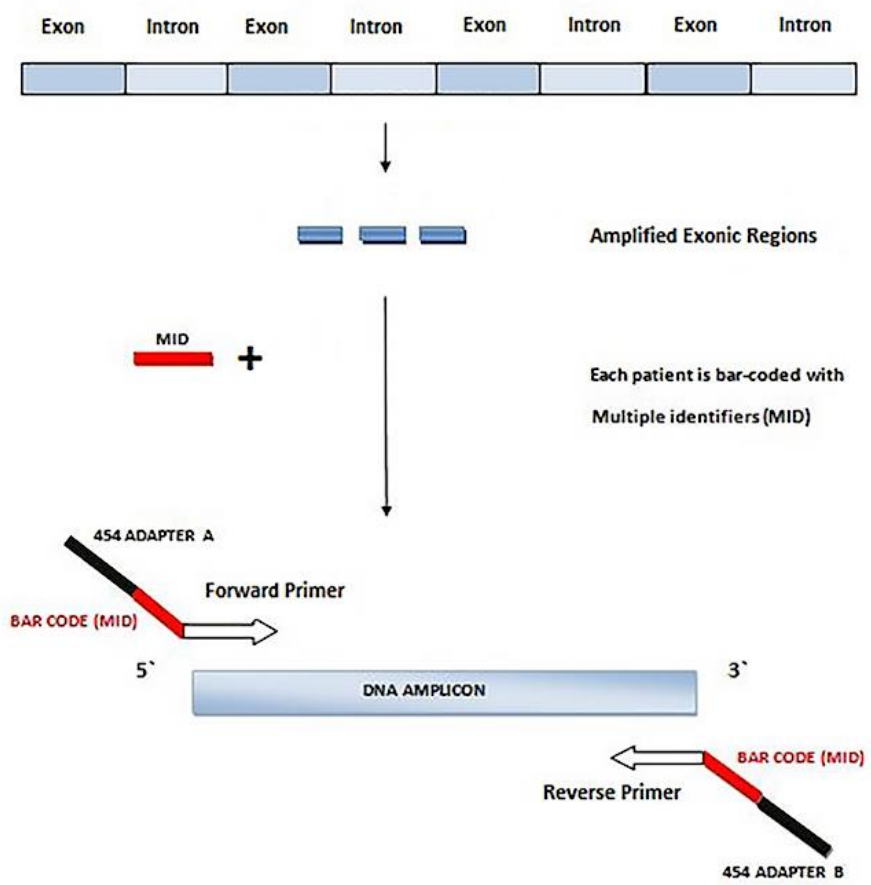


Figure 11: Multiple identifier (MID) primer design used in second round PCR.
Common bi-directional 454 specific primer sequences required for 454 sequencing are incorporated within the MIDs (A=forward; B=reverse). “MID-454 A” primers are targeted to UaS1 and “MID-454 B” primers are targeted to UaS2 in the second round PCR.

The concentration of all gene amplicons (second round PCR) for 5 randomly selected patients was measured by using the Quant-iT pico-green dsDNA assay kit as described previously in section 2.2.2. Based on the concentrations obtained from picogreen assay, all amplicons for each patient were pooled in one reaction. Following on, all patient mixes were measured by using the Quant-iT pico-green dsDNA assay kit as described previously in section2.2.2. Subsequently, upto 30 patient mixes were pooled together in equal concentrations. All pooled PCR libraries were purified using the Agencourt AMpure XP beads as described previously in section 2.2.4.2 and then quantitated using the Quant-iT picoGreen dsDNA assay kit as described previously in section 2.2.2. Pooled libraries were further processed for sequencing in a batch-wise manner. Each pooled PCR library was diluted in nuclease free H_2O till the final

concentration of 1 x10^7 molecules/µl was reached. 10µl to 12µl of diluted DNA library was subsequently used in the next step (Emulsion PCR).

2.2.5.2 Emulsion Based Clonal Amplification (emPCR) of DNA Library Amplicons

GSFLX Titanium LV emPCR Kits (Roche 454 Life Sciences) were used to perform the emulsion PCR (Figure 12) using pooled DNA library obtained from section 2.2.5.1.

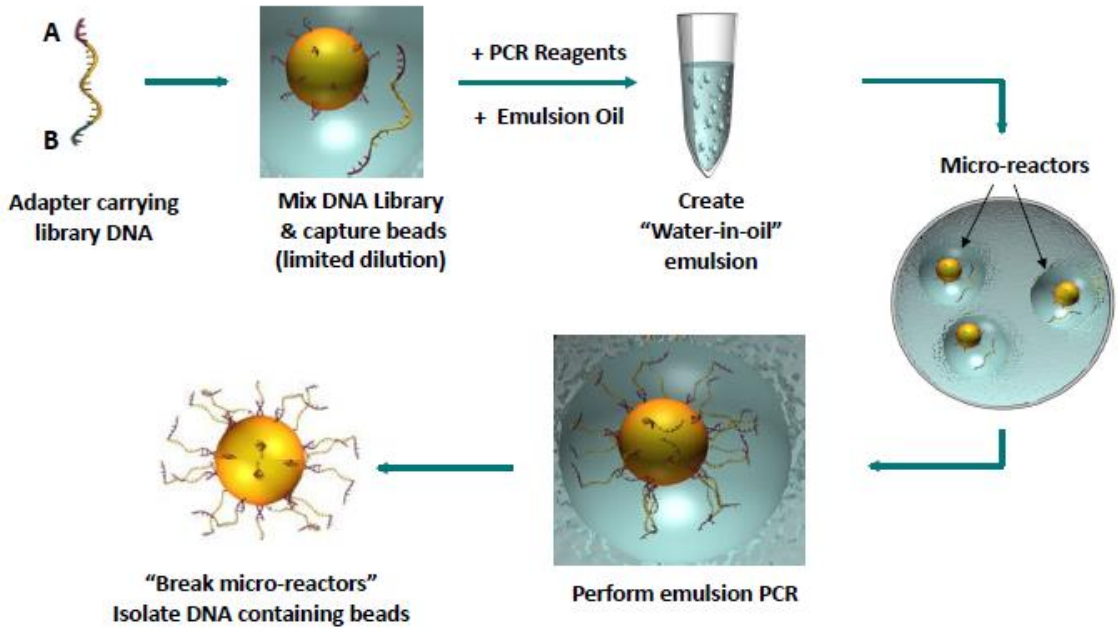


Figure 12: Schematic representation of emulsion based clonal amplification.

DNA capture beads-A (with forward primer) and -B (with reverse primer) were washed twice with 1ml of 1x capture bead wash buffer TW (Roche 454 Life Sciences). After the second wash, beads were resuspended in 100µl 1x capture bead wash buffer TW. 10µl to 12µl of the pooled DNA library (concentration 1x10^7 molecules/µl) was added to each DNA capture beads-A and -B tubes.

Two cups containing the emulsion oil/mock amplification mix were placed on the tissue lyser (Qiagen) and shaken for 5 minutes at 28 strokes/sec. 4mls of live amplification mix A/B (Appendix Table 13), DNA library A/B and DNA capture beads A/B were added to each premixed oil/mock amplification mix emulsion cup. The cups were

loaded on the tissue lyser and shaken for 5 minutes at 12 strokes/sec. The emulsion (containing DNA library and DNA capture beads) from each cup was then dispensed into the 96 well PCR plate. Four 96 well PCR plate containing the emulsified amplification reactions were loaded on the ABI 9700 and subjected to PCR (Appendix Table 14).

Emulsions containing the DNA beads with amplified DNA were processed using the GS FLX Titanium emPCR breaking kit and breaking reagents (Roche 454 Life Sciences). 100µl of isoproponol (100%) was added to each PCR plate well containing the emulsions and then the assembled vaccum/breaking manifold was used to draw the emulsion-isoproponol mix from all 4 plates into four 50ml tubes. 96 well PCR plates were washed three times with 100µl of 2-isopropanol (100%, v/v). Amplified DNA capture bead-A and -B libraries were mixed at this stage and the 50ml tubes were centrifuged at 1000xg for 10 minutes. The supernatant was discarded and the amplified DNA capture beads were washed with enhancing fluid (EF), 100% (v/v) isopropanol, 100% (v/v) ethanol and then with EF, in a stepwise manner with a centrifugation (@ 950xg) step in-between each wash. In order to denature the double stranded DNA, 1ml fresh melt solution (19.5mls of molecular grade water + 500µl of 5M NaOH) was added to the amplified DNA capture beads, followed by vigorous vortexing and centrifugation. After repeating it one more time, amplified ssDNA capture beads (single stranded DNA capture beads) were washed twice by adding 1ml of annealing buffer-XT, followed by centrifugation. Following on, amplified ssDNA capture beads were resuspended in 45µl of annealing buffer-XT. 12.5µl of enrichment primer A and 12.5µl of enrichment primer B were added to each 1.5mls tube, followed by incubation at 65°C on a heating block for 7 minutes to enable primer annealing. Following on, tubes were placed on ice for 2 minutes. Amplified ssDNA capture beads were washed 3 times by adding 1ml of ET, followed by centrifugation. Beads were resuspended in 800µl of EF. 80µl of previously washed enrichment beads were added to the amplified DNA capture beads, followed by vortexing for 1 minute. The tubes were incubated at room temperature on a lab rotator for 30 minutes. After incubation, bead mixture was transferred into a 15mls tubes and placed on the MPC for 5 minutes. The supernatant was discarded and the beads were washed 10 to 14 times with 1ml to 3mls of EF, till a clear supernatant was observed. The beads were resuspended in

1ml of melt solution (19.5ml of molecular grade water + 500µl of 5M NaOH) and vortexed for 30 seconds. The tubes were placed on the MPC for 5 minutes and the supernatant containing the enriched amplified ssDNA capture beads was transferred into a fresh 1.5ml eppendoff tube. Enriched amplified ssDNA capture beads were washed 3 times by adding 1ml of annealing buffer XT, followed by centrifugation. 200µl of annealing buffer, 25µl of sequencing primer A and 25µl of sequencing primer B was added to each tube with enriched amplified ssDNA capture beads, followed by incubation at 65°C on a heating block for 7 minutes to enable primer annealing. Following on, tubes were placed on ice for 2 minutes. ssDNA capture beads were washed 3 times by adding 1ml of annealing buffer-XT, followed by centrifugation at 3000xg for 2 minutes. The bead pellet was resuspended in 1ml of annealing buffer-XT and stored at +4°C.

2.2.5.3 Sequencing/Genome Sequencer 454/FLX setup

The sequencing was performed on Roche/454 FLX Sequencer and all the reagents were supplied by the Roche diagnostics (Roche 454 Life Sciences).

The picotiter-plate (PTP) was submerged in bead buffer2. Enriched DNA capture beads were mixed with DNA bead incubation enzyme mix and incubated for 30 minutes. DNA capture beads, packing beads, enzyme beads and PPiase beads were loaded on the PTP in a step wise manner (Figure 13). The PTP was loaded on the Roche/454 FLX sequencer and after performing the required checks on the sequencer, the sequencing run was started by using the GS sequencer software.

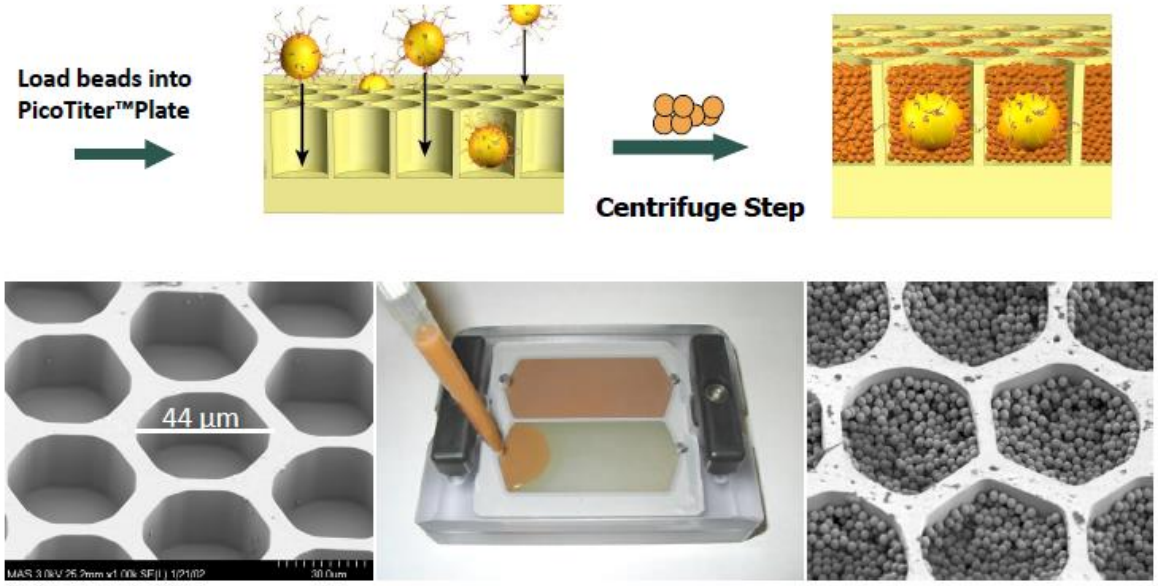


Figure 13: Depositing DNA beads into the PTP device.

One side of the PTP is polished and the other side of the plate contains wells that are 75 picoliters in volume and 44µm in diameter. Each PTP comprises 1.6 million wells. The diameter of the wells is designed so that only a single capture bead will fit into each well. Single DNA capture beads are held within the PTP wells by the packing beads. Enzyme containing beads that catalyse the downstream pyrosequencing reaction steps are also added to the PTP.

During the sequencing run, as a result of the nucleotide and other enzyme flow from the fluidics system on top of the PTP, hundreds of thousands of beads embedded within the PTP apertures with each bead carrying millions of copies of a unique sstDNA molecule are sequenced in parallel (Figure 13). The combination of signal intensity and positional information generated across the PTP is captured by the charge coupled device (CCD) camera device. These raw reads are processed by the Roche/454 FLX analysis software and then filtered by various quality filters to remove poor-quality sequences, mixed sequencing reads (generated as a result of more than one initial DNA fragment per bead), and sequences without the initiating sequence specific to the 454 adapter sequence (Figure 14).

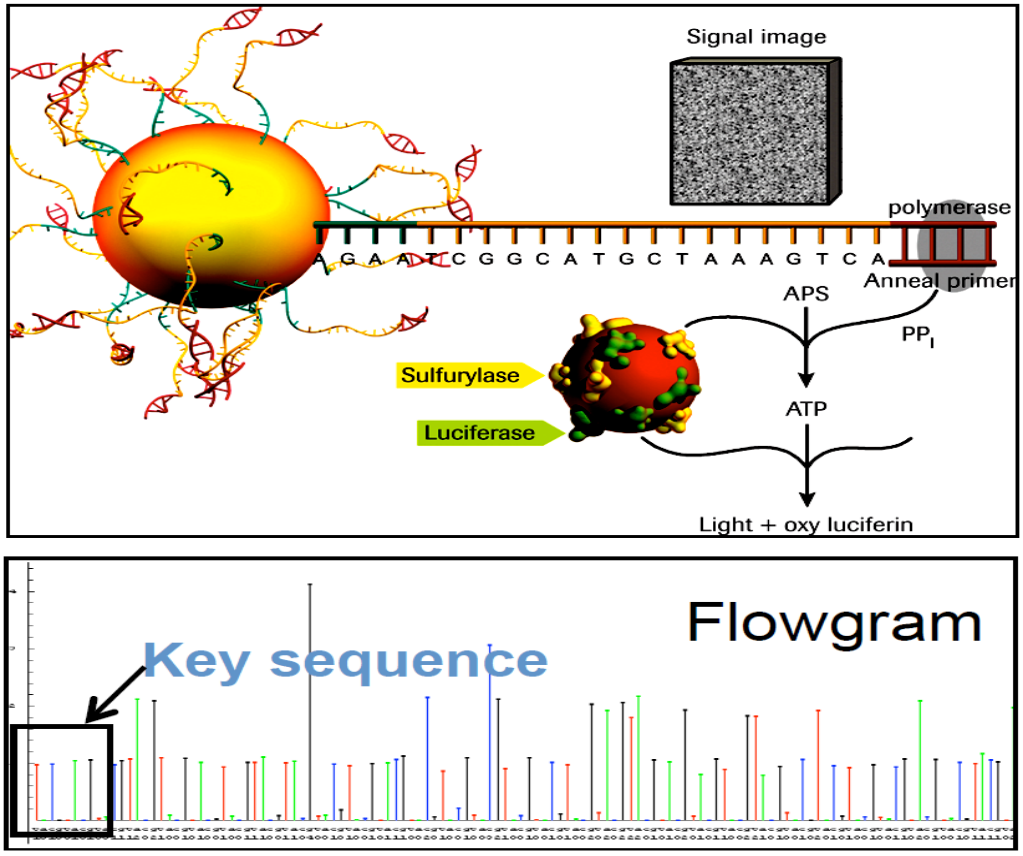


Figure 14: Roche/454 FLX Sequencing technology.

In pyrosequencing, incorporation of a nucleotide by DNA polymerase results in the release of pyrophosphate and each incorporation reaction is accompanied by release of pyrophosphate (PPi). ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5/ phosphosulfate. This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light proportional to the amount of ATP. The light generated in the luciferase-catalysed reaction is detected by a CCD camera and seen as a peak in a pyrogram. Each light signal is proportional to the number of nucleotides incorporated.

2.2.5.4 Roche targeted amplicon data analysis

Data analysis was performed using GS Amplicon variant analyser software (Roche 454 Life Sciences). An analysis project was generated by importing the reference gene sequences (Ensembl), primer sequences of all the sequenced genes and the sequences of all the MIDs used in the amplification process into the GS Amplicon variant analyser software (Roche, Germany). The data generated by the Roche/454 FLX sequencer was imported into the analysis project, followed by demultiplexing and

subsequently aligning the data to the reference sequence of the genes present on the AML gene panel. Following on, a list of variants was generated for further validation.

2.2.6 MiSeq Amplicon Sequencing

2.2.6.1 Library Preparation

PCR amplicon libraries were prepared by using the first round PCR protocol as described in section 3.2.4.1. Following on, PCR libraries were purified and quantified as described previously in section 2.2.4.2 and section 2.2.2, respectively. The product size and specificity for amplicons was also confirmed by gel electrophoresis. Patient specific PCR amplicons were normalized and mixed in batches. Following on, all patient mixes were measured by using the Quant-iT pico-green dsDNA assay kit as described previously in section 2.2.2. Pooled libraries were further processed for sequencing in a batch-wise manner. Each pooled patient amplicon libraries was diluted in nuclease free H_2O till the final concentration of 0.2 ng/µl was reached. Diluted patient amplicon libraries were subsequently used in the next step.

Following on, transposon-based Illumina Nextera technology (Illumina, UK) was used to prepare the Illumina sequencing libraries. The Nextera-XT DNA Sample Preparation Kit was used to prepare sequencing libraries by following the manufacturer's protocol (with modifications). 1.25µl (0.25ng) of each pooled patient amplicon library (0.2ng/µl) was transferred to a new 96 well PCR plate (PCR plate). 2.5µl of tagment DNA (TD) buffer and then 1.25µl of amplicon tandem mix (ATM) was added to each PCR well, respectively. Library mixtures were mixed by using multichannel pipette, followed by centrifugation @ 280xg for 1 minute. PCR plate was incubated at 55°C (5 minutes) and then 4°C (2 minutes). Following on, 1.25µl of neutralize tagment (NT) buffer was added to each mixture in the PCR plate. Samples were incubated at room temperature for 7 minutes. 3.75µl of nextera PCR master mix (NPM) was added to each well of the PCR plate. 1.25µl of index-1 primer and 1.25µl of index-2 primer was added to each well in the PCR plate. These dual indexes allow multiplexing of 96 samples independent sequencing experiments (patient specific amplicons) utilizing Nextera-amplicon methodology. Following on, mixtures were mixed by using multichannel pipette, followed by centrifugation @280xg for 1 minute. PCR amplification was performed as mentioned in Table 13. Amplified libraries were purified using Agencourt

AMpure XP beads as described in section 2.2.4.2. DNA samples were eluted in 50µl of nuclease free H_2O.

Following on, all patient mixes (post nextra amplification) were measured by using the Quant-iT pico-green dsDNA assay kit as described previously in section 2.2.2. Subsequently, patient libraries were pooled together in equal concentrations. Pooled patient libraries were purified using the Agencourt AMpure XP beads as described previously in section 2.2.4.2 and then quantitated using the Quant-iT picoGreen dsDNA assay kit as described previously in section 2.2.2. Libraries were sequenced on the Illumina MiSeq platform utilizing version 2 chemistry with 150-250 paired-end reads.

Table 13: Thermal cycler conditions used for the Nextra-PCR.

Steps	Temperature	Time
Step 1	72 °C	3 Minutes
Step 2	95 °C	30 Seconds
Step 3	95 °C	10 Seconds
Step 4	55 °C	30 Seconds
Step 5	72 °C	30 Seconds
	Repeat Step 3 through step 5 for a total of 12 cycles	
Step 5 - Final Extension	72 °C	5 Minutes
Step 6 - Hold	4 °C	hold

2.2.6.2 MiSeq data analysis

All the data analysis was performed by on-board MiSeq data analysis pipeline. Illumina RNASeq 2000 Real Time Analysis (RTA) software generates the *.bcl files during the sequencing run. *.bcl files were de-multiplexed and converted to fastqfiles. Sequence alignment to the human target reference genes (NCBI37/Hg19) and variant calling for SNPs/Indels was performed via Burrows-Wheelers aligner (BWA) (Li and Durbin, 2009) and GATK (McKenna et al., 2010) pipeline. Processed VCF and BAM files were visualized using variant studio (Illimina) and integrated genome viewer (IGV), respectively. Candidate variants obtained were passed for validation if the variants were deemed to cause protein changes and not found in dbSNP142, esp5400 and 6500, and 1000 genomes databases at >0.01 population allele frequency.

2.2.7 Sanger Sequencing

Target genes were amplified (first round PCR) as mentioned in section 2.2.5.1. Amplified PCR product was purified using Agencourt AMpure XP beads as described previously in section 2.2.4.2. Following on, cycle sequencing PCR was performed by using purified first round PCR product (5ng to 20ng), big dye enzyme master-mix, 5x sequencing buffer, (Applied Biosystems) aUS1 (forward) and aUS2 (reverse) sequencing primers (Table 14 and Table 15). The PCR product was purified with BigDye X-Terminator™ purification kit (Applied Biosystems). 45ul of SAM solution and 10µl of X-terminator solution were added to the 10µl PCR product. The plate was incubated at room temperature on a shaker for 30 minutes, followed by centrifugation at 1000xg for 2 minutes. The plate was then loaded on the ABI 3130xl Genetic Analyser and the sequencing program was set up by using the manufacturer's protocol. Data was analysed for mutations by using Applied Biosystems SeqScape® software v2.6.

Table 14: PCR reagents used to make master-mix.

Reagent	Volume per reaction
5x Sequencing buffer	2 µl
Big Dye enzyme	0.5 µl
Primer (5 µM)	1 µl
Nuclease free H_2O	as required
DNA template	15 ng
Final Volume	10 µl

Table 15: Thermal cycler conditions used for the PCR.

Steps	Temperature	Time
Step 1 - Denaturation	96 °C	1 Minutes
Step 2 - Denaturation	96 °C	10 Seconds
Step 3 - Annealing	55 °C	5 Seconds
Step 4 - Extension	60 °C	75 Seconds
Repeat Step 2 through step 4 for a total of 15 cycles		
Step 5 - Denaturation	96 °C	10 Seconds
Step 6 - Annealing	55 °C	5 Seconds
Step 7 - Extension	60 °C	90 Seconds
Repeat Step 5 through step 7 for a total of 5 cycles		
Step 8 - Denaturation	96 °C	10 Seconds
Step 9 - Annealing	55 °C	5 Seconds
Step 10 - Extension	60 °C	120 Seconds
Repeat Step 8 through step 10 for a total of 5 cycles		
Step 12 Hold	10 °C	Infinite

2.2.8 Isolation of human haematopoietic stem cells and mature progenitors

2.2.8.1 Mononuclear cell separation

Mononuclear cells (MNCs) were separated from patient bone marrow sample by density gradient sedimentation with Histopaque-1077 (Sigma). Briefly, bone marrow sample was transferred into a 50ml tube containing the same volume of Histopaque-1077 (1:1 ratio). Bone marrow cells were centrifuged at 450xg for 20 minutes. MNCs were harvested from the histopaque/plasma interface and transferred into a new 5ml round-bottom tube. MNCs were washed twice with AutoMACS running buffer (Miltenyl Biotec), centrifuged at 300xg for 10 minutes and finally counted. Up to $1x10^8$ MNCs were resuspended in 350µl of auto MACS running buffer and subsequently used for immunomagnetic enrichment of $CD34^+$ and $CD34^-$ cells using magnetic activated cell sorting technology (Miltenyl Biotec).

2.2.8.2 $CD34^+$ and $CD34^-$ cell selection by Magnetic activated cell sorting (MACS)

First, the $CD34^+$ cells were magnetically labelled with CD34 MicroBeads (Miltenyl Biotec). 100µl of FcR blocking reagent (Miltenyi Biotec) and 50µl of CD34 micro beads (Miltenyi Biotec) were added to MNCs (up to 1 x 10^8 cells), followed by vortexing. MNCs were incubated for 15 minutes at 4ºC. 3ml of AutoMACS running buffer was added to CD34 bead cell suspension. The magnetically labelled CD34 cell suspension was loaded onto a MACSR column in the MACS separator. Posseld program was used to isolate $CD34^+$ cells from MNCs suspension. The magnetically labelled $CD34^+$ cells are retained within the MACS column. The unlabelled cells run through; this cell fraction is thus depleted of $CD34^+$ cells and only contained $CD34^-$ cells. After removing the column from the magnetic field, the magnetically retained $CD34^+$ cells can be eluted as the positively selected cell fraction. The Posseld program increases the purity of $CD34^+$ cells by enriching the $CD34^+$ cells twice during the separation process. Both cell populations ($CD34^+$ and $CD34^-$ cells) were washed twice by adding 1ml of FACS wash buffer (PBS, 2% FCS) , followed by centrifugation at 300xg for 10 minutes and resuspended in 1ml of FACS wash buffer. Cells were counted and then stored on ice. $CD34^+$ cells were centrifuged at 300xg for 10 minutes and then supernatant was discarded. $CD34^+$ cells were either processed for haematopoietic stem cell separation using FACS or stored at -20ºC. Previous studies have reported that a purity of ≥94% can be achieved with this isolation method (Bonanno et al., 2009; Giarratana et al., 2005).

$CD34^-$ cells were divided into two aliquots. One aliquot was processed for FACS experiment and another aliquot was used for enrichment of $CD235^+$ and $CD235^-$ cells using MACS. 10x10^6 $CD34^-$ cells were used for MACS separation. Cells were centrifuged at 300xg for 10 minutes and then supernatant was discarded. 20µl of CD3 micro beads (Miltenyl Biotec)and 80µl of FACS wash buffer was added to $CD34^-$ cells (up to upto 1 x 10^7 cells), followed by vortexing and incubating for 15 minutes at 4ºC. 3ml of AutoMACS running buffer was added to CD3 cell suspension. The magnetically labelled $CD3^+$ cell suspension was loaded onto a MACS column in the MACS separator. Possel program was used to isolate $CD3^+$ cells from $CD34^-$ cell suspension. $CD34^-CD3^-$ cells were centrifuged at 300xg for 10 minutes and then supernatant was

discarded. 20µl of CD235a micro beads (Miltenyl Biotec) and 30µl of FACS wash buffer was added to the $CD34^{-}CD3^{-}$ cells, followed by vortexing and incubating for 15 minutes at 4ºC. 1ml of AutoMACS running buffer was added to $CD34^{-}CD3^{-}$ cell suspension. The magnetically labelled cell suspension was loaded onto a MACS column in the MACS separator. Possel program was used to isolate $CD34^{-}CD3^{-}CD235^{+}$ and $CD34^{-}CD3^{-}CD235^{-}$ cells from $CD34^{-}CD3^{-}$ cell suspension. $CD34^{-}CD3^{-}CD235^{-}$ cells were centrifuged at 300xg for 10 minutes, supernatant was discarded and cells were stored as viable cells at -20ºC.

2.2.8.3 Fluorescent activated cell sorting cell sorting (FACS) - Antibodies, Staining and cell sorting

MACS purified $CD34^{+}$ and $CD34^{-}$ cells were washed twice by adding 1ml of FACS wash buffer, followed by centrifugation at 300xg for 10 minutes. Both cell fractions were stained either with antibodies present in antibody panel A (Table 16) or antibody panel B (Table 17) or antibody panel C (Table 18). Appropriate amount of PBS was added to the tubes to make a final volume of 100µl. Cells and antibodies were incubated for 30 minutes in the dark at 4ºC. Cells were then washed twice by adding 1ml of FACS wash buffer, followed by centrifugation at 300xg for 10 minutes. Supernatant was discarded and cells were resuspended in PBS @4ºC at $5x10^{6}$ cells/ml. Propidium iodide (PI) was added to the cell suspension prior to sorting to exclude dead cells. Cells were then placed on ice and were analysed within 1 hour on BD FACS Aria I.

Table 16: Tables showing antibodies and the flourchromes used for FACS analysis.

Antibody Panel A				
Cell Surface marker	**Flourochrome**	**Catalogue number**	**Manufacturer**	**Volume (up to 10^8 cells)**
Anti-Human CD135 (Flt3)	PE	12-1357-41	eBioscience	5µl
Anti-Human CD45	APC-eF® 780	47-0458-41	eBioscience	5µl
Anti-Human CD38	APC	17-0389-41	eBioscience	5µl
Anti-Human CD34	PE-Cy7	25-0349-41	eBioscience	5µl
Anti-human CD45RA	BV510™	304141	Biolegend	5µl/10^6
Anti-human CD90	AF700	328119	Biolegend	2µl/10^6
Anti-Human/Mouse CD49f (Integrin alpha 6)	FITC	11-0495-80	eBioscience	1µl
Negative Lineage Antibodies				
Anti-Human CD2	Biotin	13-0029-80	eBioscience	1µl
Anti-Human CD3	Biotin	13-0037-80	eBioscience	1µl
Anti-Human CD14	Biotin	13-0149-80	eBioscience	2µl
Anti-Human CD16	Biotin	13-0168-80	eBioscience	1µl
Anti-Human CD19	Biotin	13-0199-80	eBioscience	1µl
Anti-Human CD24	Biotin	13-0247-80	eBioscience	1µl
Anti-Human CD56	Biotin	13-0567-80	eBioscience	1µl
Anti-Human CD235a	Biotin	13-9987-80	eBioscience	1µl
Anti-Human CD66b	V450	561649	BD	5µl
Streptavidin	EF® 450	48-4317-82	eBioscience	1µl

Table 17: Tables showing antibodies and the flourchromes used for FACS analysis.

Antibody Panel B				
Cell Surface marker	**Flourochrome**	**Catalogue number**	**Manufacturer**	**Volume (up to 10^8 cells)**
Anti-Human CD34	APC	unknown	eBioscience	5µl
Anti-Human CD71	PE	unknown	eBioscience	5µl
Anti-Human CD235	FITC	unknown	eBioscience	5µl
Negative Lineage Antibodies				
Anti-Human CD2	Biotin	13-0029-80	eBioscience	1µl
Anti-Human CD3	Biotin	13-0037-80	eBioscience	1µl
Anti-Human CD14	Biotin	13-0149-80	eBioscience	2µl
Anti-Human CD16	Biotin	13-0168-80	eBioscience	1µl
Anti-Human CD19	Biotin	13-0199-80	eBioscience	1µl
Anti-Human CD24	Biotin	13-0247-80	eBioscience	1µl
Anti-Human CD56	Biotin	13-0567-80	eBioscience	1µl
Anti-Human CD66b	V450	561649	BD	5µl
Streptavidin	eFluor® 450	48-4317-82	eBioscience	1µl

Table 18: Tables showing antibodies and the flourchromes used for FACS analysis.

Antibody Panel C			
Cell Surface marker	**Flourochrome**	**Manufacturer**	**Volume (up to 10^8 cells)**
Anti-Human CD34	PE-Cy7	eBioscience	5 µl
Anti-Human CD3	PE	BD	5 µl
Anti-Human CD4	EF450	BD	5 µl
Anti-Human CD19	APC	BD	5 µl

The BD FACS Aria I (San Jose, CA) system is a four-color sorter, with 5 lasers and 15 detectors capable of both cell analysis and cell sorting. Fluorescently labelled cells were then analysed using dot plots measuring forward versus side scatter. Using the appropriate flow cytometry values, our cell populations of interest within the stem cell compartment were defined by hierarchical gating which was largely similar to the one used by *Notta et al* (Notta et al., 2011); HSCs as $Lin^-CD45^+CD34^+CD38^-CD45RA^-CD90^+CD49f^+$, MPPs as $Lin^-CD45^+CD34^+CD38^-CD45RA^-CD90^-CD49f^-$, GMPs as $Lin^-CD45^+CD34^+CD38^+CD135^+$ $CD45RA^+$ and MEPs as $Lin^-CD45^+CD34^+CD38^+CD135^-CD45RA^-$ (Figure 15). In addition, mature erythoid and lymphoid cells were also isolated by gating for specific cell surface markers such as $Lin^-CD34^-CD235^+CD71^+$, $Lin^-CD34^-CD235^+CD71^-$, $Lin^-CD34^-CD71^+$, $Lin^-CD34^-CD235^+$ (Figure 16), $CD34^-CD3^+$, $CD34^-CD3^+4^+$ and $CD34^-CD19^+$ (Figure 17). After sorting, the purity of all the cell fractions was checked and a sorting purity of ≥97% was routinely obtained. FACS sorted cells were centrifuged at 450xg for 10 minutes and then supernatant was discarded. Cells were stored as dry pellet or in trizol at -20ºC.

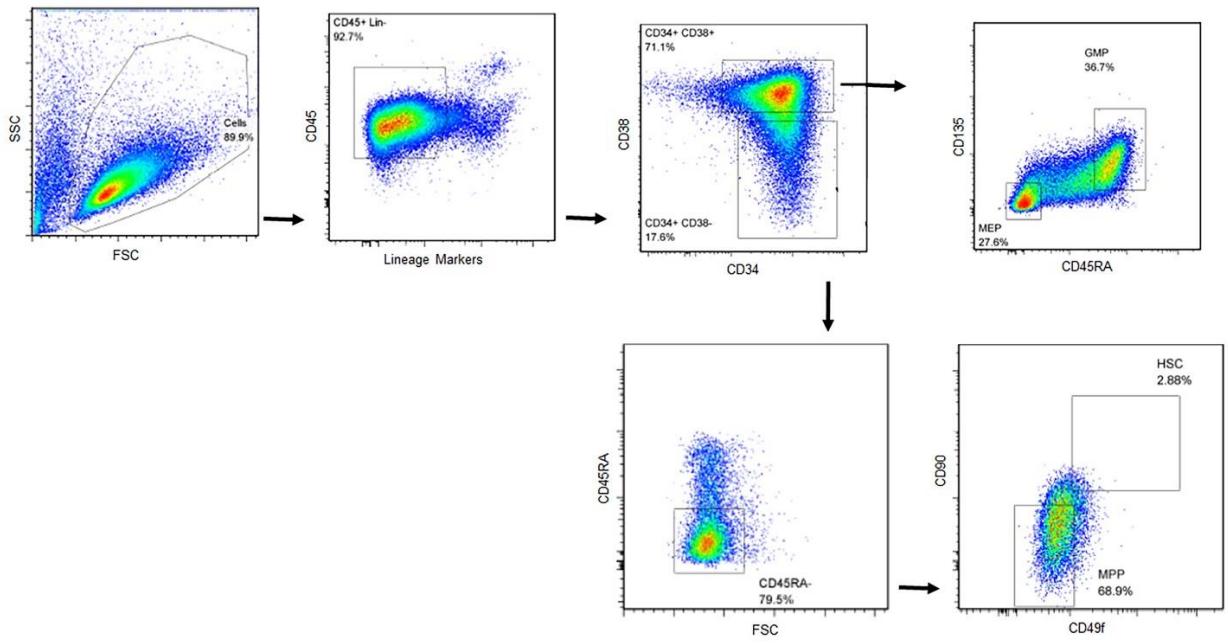


Figure 15: Representation of the isolated hematopoietic stem cell subpopulations.
Viable lineage negative $CD34^+$ cells were defined by hierarchical gating on a combination of forward (FSC) and side scatter (SSC), propidium iodide (not shown), CD45 and a lineage cocktail. These $CD34^+$ cells were further gated first into $CD34^+CD38^-$ and $CD34^+CD38^+$ populations and then subdivided into either HSCs ($CD34^+CD38^-CD90^+CD45RA^-CD49f^+$) and MPPs ($CD34^+CD38^-CD90^-CD45RA^-CD49f^-$) or into GMPs ($CD34^+CD38^+CD135^+CD45RA^+$) and MEPs ($CD34^+CD38^+CD135^-CD45RA$).

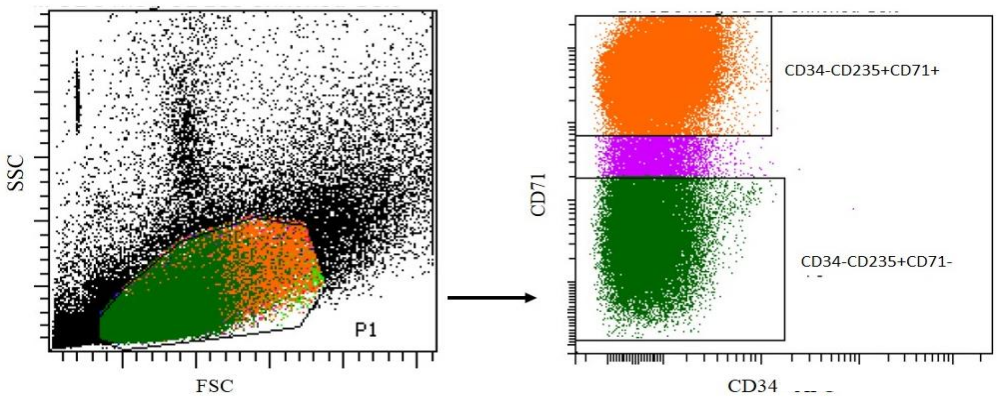


Figure 16: Representation of the isolated cell populations.
$CD34^-CD3^-CD235^+CD71^+$ and $CD34^-CD3^-CD235^+CD71^-$ cells were defined by gating on a combination of forward (FSC) and side scatter (SSC), propidium iodide (not shown).

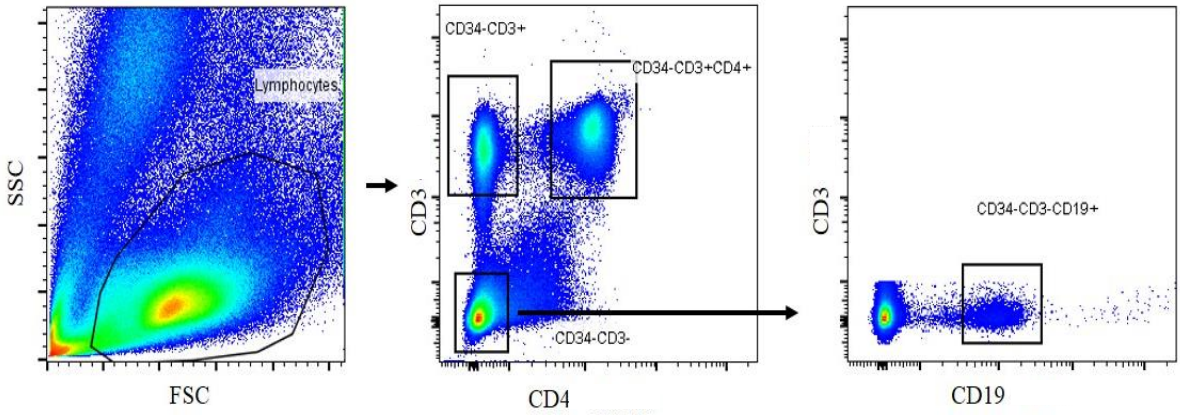


Figure 17: Representation of the isolated cell populations.

$CD34^{-}CD3^{+}$, $CD34^{-}CD3^{+}CD4^{+}$ and $CD34^{-}CD3^{-}CD4^{-}CD19^{+}$ were defined by gating on a combination of forward (FSC) and side scatter (SSC), propidium iodide (not shown).

2.2.9 Statistical analysis

The characteristics of the study population were studied with appropriate statistical methods (Mann-Whitney test for continuous variables and Fisher's exact test for categorical variables) comparing patients with specific mutation versus cases without mutation. Clinical characteristics, survival and time to progression to AML were updated to January 2012 and measured from time of sample collection (N=123) or diagnosis (N=31) at Kings College Hospital. The median disease duration prior to sample analysis for 123 patients was 11 months (range 1-119 months), during this period 29 patients showed disease progression (upstaging of WHO category) of which 10 were treated with either with intensive chemotherapy or 5-azacitidine.

The Kaplan-Meier estimate was used to evaluate time to survival and time to progression. The log-rank test was used to assess potential differences in outcome between subgroups. A p-value of ≤0.05 was considered statistically significant.

Prism Version 6 software (GraphPad) was also used for statistical analysis. Data is presented as the mean ± S.E.M. Statistical analysis was performed using the two-tailed student's t-test for comparison of two groups to determine the level of significance.

2.2.10 NSG Xenotransplant Assay

MNCs were isolated from the bone marrow cells by centrifugation using Ficoll-Paque™ PLUS (GE Healthcare Life Sciences, Buckinghamshire, UK). The cells were processed within 24 hours following collection using an Easysep Human CD34 positive selection kit and Easysep magnet (StemCell Technologies, Vancouver, Canada) according to the manufacturer's instructions. 1 ml of RoboSep buffer was added to MNCs (up to 5 x 10^8 cells). Following on, 100µl of EasySep® Positive Selection Cocktail was added to the MNCs. Cell mixture was mixed and incubated for 15 minutes at room temperature. 50µl of EasySep® magnetic nanoparticles was added to the MNCs and mixture was incubated for 10 minutes at room temperature. After incubation, tube containing the MNC/bead mixture was placed on the magnet for 5 minutes. The supernatant was discarded and the beads were washed 4 times with 1ml of buffer. Following on, enriched $CD34^+$ cells were resuspended PBS.

NSG mice were a kind gift of Dr Leonard Shultz (The Jackson Laboratory). All animal experiments were performed in accordance to Home Office and CRUK guidelines. Prior to transplantation, mice received a sub-lethal dose of radiation (330-375 cGy) from a cesium-137 source. Direct intra–bone marrow injection was performed in the mouse tibia with $0.65x10^5$ to $2x10^5$ human bone marrow $CD34^+$ cells from patients and healthy donors.

Engraftment was assessed over time by intra-tibia aspiration under isoflurane anaesthesia and the bone marrow was immunophenotyped by the presence of $mCD45^-$, $hCD45^+$, $CD33^+$, $CD19^+$ and $CD3^+$ (using anti-mouse and anti-human antibodies from BD Biosciences, Oxford, UK) cell populations. After 18-20 weeks, mice were terminated and cells were harvested from pooled bone marrow (femurs, tibias, pelvis and/or spine). Live cells were stained and sorted on $hCD45^+$ phenotype using FACS Aria SORP (BD Biosciences, Oxford, UK). Sorted cells were washed in PBS and harvested in order to later perform genomic analysis. All the xenograft experiments were performed by Dr Kevin Rouault-Pierre.

2.2.11 Single cell clonogenic assay

Five hundred cells from the bone marrow $CD34^+$enriched fraction were plated in triplicate in 0.5 mL of MethoCult H4434 (StemCell Technologies, Vancouver, Canada)

in 24 wells tissue culture dishes to assess colony-forming units. At day 14 of culture, the numbers of colonies were scored. Cells were picked, harvested and washed twice with PBS before analysis. Single cell colony images were acquired with a Zeiss AxioVert 40 CFL microscope using 5X lens and connected to Canon PowerShot A640 digital camera. Single colonies were picked using a Zeiss AxioVert 40 CFL microscope, 10μl tips were used to aspirate the colonies which were then resuspended in 10μl of PBS in 96 well plate maintain on ice.

All colonies were subjected to whole-genome amplification (Single cell WGA, Sigma-Aldrich) as described in section 2.2.13. All colony forming cell (CFC) colonies were screened for *SF3B1* and additional patient specific point mutations. Amplicon libraries were amplified through PCR using the primers specific to gene mutations and subjected to Nextra-Illumina sequencing as described in section 2.2.6. Gene mutation for each single colony was called when >30% of the sequencing reads were carrying a mutant allele.

2.2.12 Long-term colony Assay

LTC assays were performed by plating 10^3 $CD34^+$ bone marrow cells in quadruplets on irradiated MS-5 murine stromal cells and cultured in Myelocult H5100 (StemCell Technologies, Vancouver, Canada) in the presence of cytokines (20 ng/ml G-CSF, 20ng/ml IL-3, and 20 ng/ml TPO from PeproTech, London, UK). After 5 weeks, live cells were stained and sorted on human $CD45^+$ phenotype using the FACS Aria SORP (BD Biosciences, Oxford, UK). Sorted cells were washed in PBS, harvested and stored as cell pellet and/or resuspended in Karnoy's solution in order to later perform FISH analysis. These LTC experiments were performed along with Dr Kevin Rouault-Pierre.

2.2.13 Whole-genome Amplification

Whole Genome amplification (WGA) of cells was performed with GenomePlex Single Cell Whole genome amplification kit (WGA4, Sigma-Aldrich Co, LLC) and/or GenomePlex SeqPlex XE kit (SEQXE, Sigma-Aldrich Co, LLC). WGA amplification was performed according to the instructions of the manufacturer, along with a no cell reaction as a negative control and a reaction of human tissue genomic DNA as positive control. For patient samples where only one mouse was used in *in-vivo* experiments,

two independent amplification experiments were performed using two different amplification kits (i.e. Sigma WGA4 and Sigma SEQXE). In addition, HSCs and MPPs for all patient samples were also amplified by two different amplification kits (Sigma WGA4 and Sigma SEQXE) while as GMPs and MEPs were amplified by using Sigma SEQXE WGA kit.

2.2.13.1 Whole genome amplification (WGA4)

Cells were diluted in 9µl of nuclease free H_2O. 1x Lysis and fragmentation buffer solution was prepared by adding proteinase-K solution and 10x single cell lysis and fragmentation buffer (ratio 1:8). 1µl of the freshly prepared proteinase-K solution-10x single cell lysis and fragmentation buffer was added to the cells. Samples were mixed and incubated in a thermal cycler at 50°C for 1 hour, then heated to 99°C for four minutes. Following on, cells were kept on ice for 5 minutes and then subjected to centrifugation at 300xg for 1 minute. 2µl of 1x single cell library preparation buffer and 1µl of library stabilization solution was added to each well containing the cells. Samples were mixed thoroughly, centrifuged (300xg for 1 minute) and place in thermal cycler at 95°C for 2 minutes, then cooled to 4°C for 2 minutes. 1µl of library preparation enzyme was added to the cells. Cells were incubated in the thermal cycler at the following temperature; 16°C for 20 minutes, 24°C for 20 minutes, 37°C for 20 minutes, 75°C for 5 minutes and then cooled to 4°C for 5 minutes. Following on, amplification master mix (7.5µl of 10x amplification master mix, 48.5µl of nuclease free H_2O and 5.0µl of WGA DNA polymerase) was added to each well containing the cells. Samples were mixed thoroughly, centrifuged briefly, and incubated in thermos cycler (Table 19). After cycling was complete, WGA samples were purified as described in section 2.2.4.2 and then stored at –20°C until ready for later analysis.

Table 19 : Thermocycler conditions used for PCR amplification

Steps	Temperature	Time
Step 1 - Denaturation	95 °C	3 Minutes
Step 2 - Denaturation	94 °C	30 Seconds
Step 3 – Annealing/ Extension	65 °C	5 minutes
	Repeat Step 2 through step 3 for a total of 25 cycles	
Step 4 – Hold	4 °C	α

2.2.13.2 Whole genome amplification (SEQXE)

Cells were diluted in 9µl of nuclease free H_2O. 1x lysis and fragmentation buffer solution was prepared by adding proteinase-K solution (WGA4) and 10x single cell lysis and fragmentation buffer (WGA4, ratio 1:8). 1µl of the freshly prepared proteinase-K solution-10x single cell lysis and fragmentation buffer was added to the cells. Samples were mixed and incubated in a thermal cycler at 50°C for 1 hour, then heated to 99°C for four minutes. Following on, cells were kept on ice for 5 minutes and then subjected to centrifugation at 300xg for 1 minute. 2µl of 1x library preparation buffer and 1µl of nuclease free H_2O was added to each well containing the cells. Samples were mixed thoroughly, centrifuged (300xg for 1 minute) and place in thermal cycler at 95°C for 2 minutes, then cooled to 4°C for 2 minutes. 1µl of library preparation enzyme was added to the cells. Cells were incubated in the thermal cycler at the following temperature; 16°C for 20 minutes, 24°C for 20 minutes, 37°C for 20 minutes, 75°C for 5 minutes and then cooled to 4°C for 5 minutes. Following on, amplification master mix (15µl of 10x Amplification Master Mix, 42.5µl of nuclease free H_2O and 1.5µl of WGA DNA Polymerase) was added to each well containing the cells. Samples were mixed thoroughly, centrifuged briefly, and incubated in thermos cycler (Table 20). After cycling was complete, WGA samples were purified as described in section 2.2.4.2 and then stored at –20°C until ready for later analysis.

Table 20: Thermocycler conditions used for PCR amplification.

Steps	Temperature	Time
Step 1 - Denaturation	94 °C	2 Minutes
Step 2 - Denaturation	94 °C	15 Seconds
Step 3 – Annealing/Extension	70 °C	5 minutes
	Repeat Step 2 through step 3 for a total of 20 cycles	
Step4 - Extension	70 °C	30 minutes
Step 5 – Hold	4 °C	α

2.2.14 Fluorescence in situ Hybridization (FISH)

Human cells were pelleted, washed twice in PBS and then resuspended in 0.5ml to 1ml of Karnoy's solution (3:1 Methanol/Acetic Acid Glacial, purchased from Sigma-Aldrich and Fisher Chemical respectively) and then stored at 4ºC prior to hybridisation. For interphase FISH, fixed cells were centrifuged and appropriate amount of supernatant was removed. Pelleted cells were then transferred onto the slides and hybridized with a XL 7q22/7q36 (Metasystems) locus-specific probe which detects deletions in the long arm of chromosome 7 (Figure 18). The orange labelled probe targets a specific region at 7q22 and the green labelled probe binds specifically to 7q36. In addition, a blue (aqua) labelled probe which hybridizes to the centromere of chromosome 7 functions as a reference probe. Fluorescence images were obtained with the help of fluorescence microscopy. 200 to 600 interphase cells were analysed per sample.

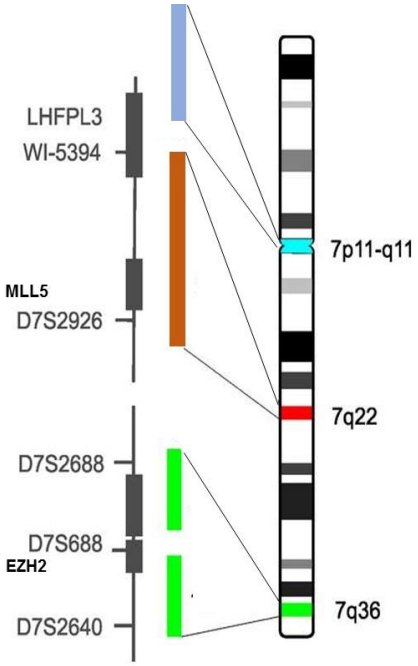


Figure 18: Detailed map of locus specific probe targeting q arm of chromosome 7.
XL 7q22/7q36 probe detects a specific region at 7q22 (orange) including the MLL5 gene while as the green labelled probe hybridizes specifically to 7q36 (green) and includes the EZH2 gene. In addition, a blue (aqua) labeled probe which hybridizes to the centromeric region of chromosome 7 acts as a reference probe.

2.2.15 Affymetrix Human Transcriptome 2.0 gene expression profiling

Total RNA was extracted from FACS purified cells as described in section 2.2.3. Purified RNA was quantified using ND-8000 (Nanodrop) spectrophotometer. Patient libraries were prepared by following manufacturer's protocol. In-between 50ng to 150ng of RNA was used for library preparation as recommended by the manufacturer. The workflow for the library preparation is depicted in Figure 19.

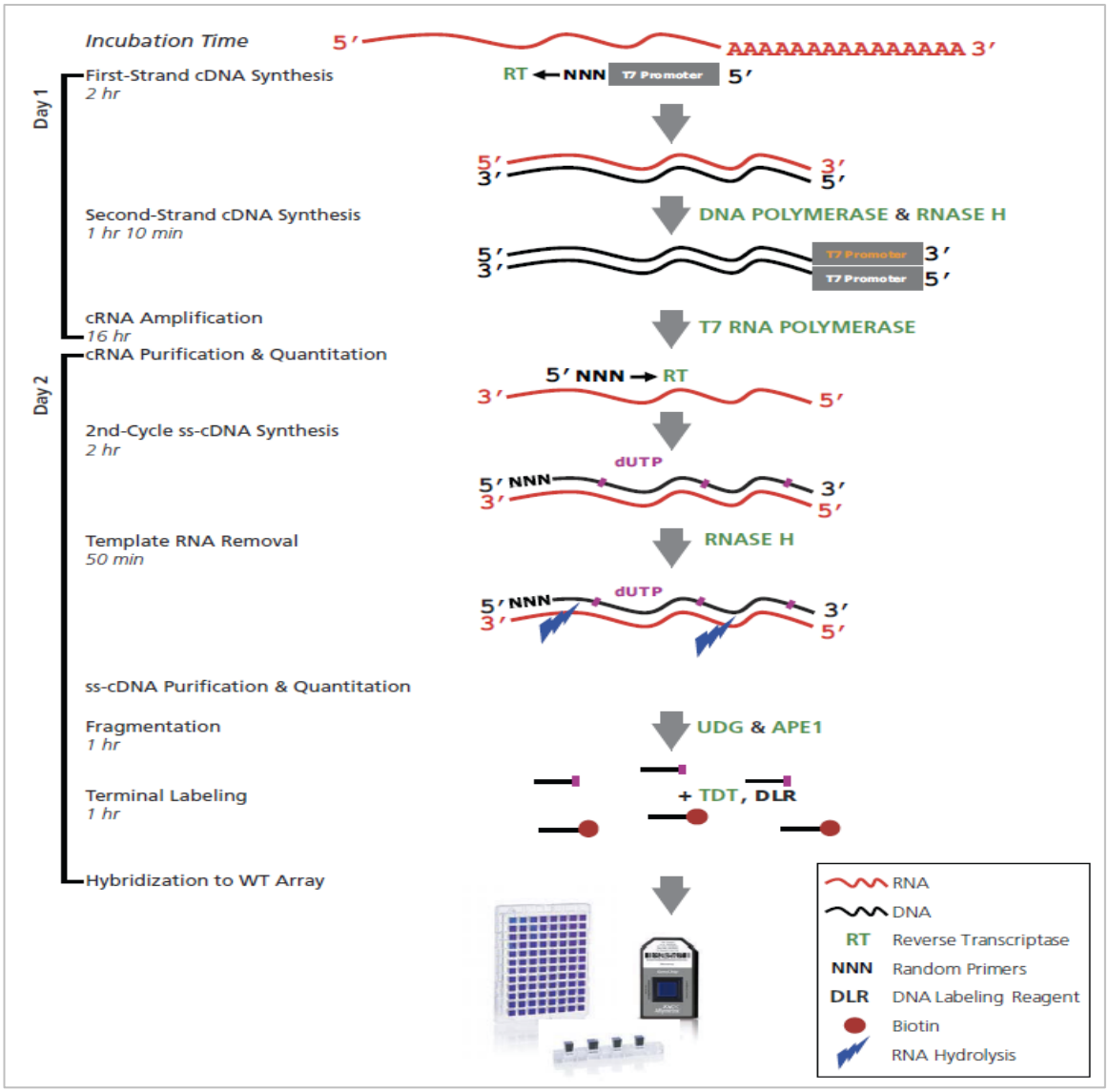


Figure 19: Work-flow for library preparation of Affymetrix HTA 2.0 array.

All reagents for library preparation are supplied by Affymetrix. cDNA, followed by cRNA is synthesized using 50ng to 150ng of purified RNA. Following on, cRNA is purified and then used to generate ss-cDNA. Ss-cDNA is purified, labelled and then hybridized to the Affymetrix chip.

2.2.15.1 Reverse transcription to synthesise first-strand cDNA

All reagents were included in the WT amplification kit. First-Strand master mix containing the 4µl of first-Strand Buffer and 1µl first-Strand Enzyme (per reaction) was prepared on ice. 5µl of the first-strand master mix was then added to 5µl of total RNA (upto 150ng). Poly-A controls at this step was not included in the reaction in order to increase the volume of patient total RNA (5µl instead of 3µl, as recommended by the

manufacturer). The reaction mixtures were mixed well and incubated in a thermocycler by using the recommended PCR program (25°C for 60 minutes, 42°C for 60 minutes and 4°C for 2 min; heated lid at 42°C).

2.2.15.2 Second-strand cDNA synthesis

Amplified libraries were immediately processed for second strand cDNA synthesis. Second-strand master mix containing the 18µl of second-strand buffer and 2µl second-strand enzyme (per reaction) was prepared on ice. 20µl of the second-strand master mix was added to the 10µl of the amplified first-strand reaction (final reaction volume 30 µl). After mixing and brief centrifugation, the samples were incubated in a thermocycler using the recommended program (16^{o}C for 60 minutes, 65^{o}C for 10 minutes and 4°C for 2 minutes).

2.2.15.3 *In-Vitro* Transcription (IVT) to synthesise cRNA

Amplified libraries were left at room temperature for 5 minutes and then immediately prepared for *in vitro* transcription. IVT master mix containing the 24µl of IVT buffer and 6µl IVT enzyme (per reaction) was prepared on ice. 30µl of the IVT master mix was added to 30µl amplified second-strand cDNA libraries (final reaction volume 60 µl). Reactions were vortexed briefly and incubated for 40°C for 16 hours with a heated lid at 40^{o}C. The samples were placed on ice and immediately prepared for purification.

2.2.15.4 cRNA Purification

The cRNA elution buffer (Nuclease-free d$H_2$0) was pre-heated to 65^{o} C for at least 10 minutes. The purification beads were mixed thoroughly by vortexing. 100µl of the beads were added to each 60µl of the cRNA reaction and transferred to a 96 well plate. The reactions were mixed and incubated at room temperature for 10 min. The cRNA samples bind to the beads during the incubation. The plate was transferred to a 96 well magnetic stand (Agentcourt, Bioscience) to capture the beads. Following on, the supernatant was aspirated and discarded. Beads (with attached cRNA) were washed three times with 200µl of 80% (v/v) ethanol. The beads were air-dried on the magnetic stand for up to 10min and the cRNA was eluted in 27µl of preheated (at

65^{o}C) nuclease-free water. Purified cRNA libraries were quantified by ND-8000 (Nanodrop) spectrophotometer.

2.2.15.5 2nd Cycle single-strand cDNA synthesis

24µl of cRNA library for each patient with a concentration of 625ng/µl (total 15µg cRNA) was transferred to a 96 well place kept on ice. 4µl of 2nd cycle primers were added to each patient library and mixed thoroughly. Libraries were incubated in a thermocycler using the recommended program (70^{o} C for 5 minutes, 25^{o} C for 5 minutes and 4^{o}C for 2 minutes) with heated lid (at 70^{o}C). After annealing the primers, the reactions were mixed briefly and centrifuged.

Mastermix containing the 8µl of ss-cDNA buffer and 4µl ss-cDNA enzyme (per reaction) was prepared on ice. 12µl of the master mix was added to each cRNA/2nd-Cycle primer reaction (final volume 40µl). After mixing briefly and centrifugation, libraries were incubated in a thermocycler using the recommended program (25^{o}C for 10 minutes, 42^{o}C then 90 minutes, 70^{o}C for 10 minutes and 4^{o}C for 2 minutes) with a heated lid (@70^{o}C).The samples were placed immediately on ice and prepared for next step.

2.2.15.6 Hydrolyse RNA by RNase H

4µl of RNase H was added to each 2nd-cycle ss-cDNA library (final reaction volume 44µl). The reactions were incubated in a thermocycler using the recommended program (37^{o}C for 45 minutes, 95^{o}C for 5 minutes and 4^{o}C for 2 minutes) with heated lid (@70^{o}C). Following on, 11µl of nuclease-free water was added to each reaction, mixed briefly and immediately prepared for purification.

2.2.15.7 Purification of 2nd- Cycle single-strand cDNA

Purification beads were mixed thoroughly. 100µl of the beads were added to 55µl of the 2nd-cycle ss-cDNA sample and transferred to a 96 well plate. The samples were mixed with 150µl of 100% ethanol by pipetting and incubated for 20 minutes at room temperature. The plate was placed on a magnetic stand for 5 minutes to capture the beads. The supernatant was carefully aspirated and discarded. Beads were washed three times with 200µl of 80% (v/v) ethanol. The beads were air-dried on the magnetic

stand for 5 minutes. cRNA libraries were eluted in 30µl of preheated nuclease-free water (@ 65ºC). Purified libraries were quantified by ND-8000 (Nanodrop) spectrophotometer.

2.2.15.8 Fragmentation and labelling of single-strand cDNA

31.2µl of the ss-cDNA library for each patient with a concentration of 176 ng/µl (total 5.5µg) was mixed with 16.8µl of fragmentation master mix which was prepared on ice (Table 21)

Table 21: Reagents used to make master-mix.

Reagent	Volume/reaction
Nuclease-free dH_20	10µl
10X cDNA fragmentation buffer	4.8µl
UDG,10 U/µl	1µl
APE 1, 1,000 U/µl	1µl
Total Volume	16.8µl

The samples were mixed gently and incubated in a thermocycler using the recommended program (37ºC for 60 minutes, 93ºC for 2 minutes and 4ºC for 2 minutes) with heated lid (@100ºC). Following on, 45µl of the fragmented ss-cDNA was transferred to new 96 well plate. 15µl of the labelling master mix (Table 22) was added to each fragmented ss-cDNA library.

Reactions were mixed, briefly centrifuged and incubated in a thermocycler using the recommended program (37ºC for 60 minutes, 70ºC for 10 minutes and 4ºC for 2 minutes). Following on, samples were briefly centrifuged and stored at -20ºC.

Table 22: Reagents used to make the labelling master mix.

Reagent	Volume/reaction
5X TdT Buffer	12µl
DNA labelling reagent, 5mM	1 µl
TdT, 30 U/µl	2µl
Total Volume	15µl

2.2.15.9 WT Array Hybridization

Dimethyl sulfoxide (DMSO) was thawed prior to use. Pre-prepared 20x hybridization controls were denatured by heating for 5 min at 65ºC. Hybridization master mix was prepared at room temperate (Table 23).

Table 23: Reagents used to make the labelling master mix.

Reagent	Volume/reaction
Control Oligo B2 (3 nM)	3.7 µl
20x Hybridization Controls *(bioB, bioC, bioD,cre)*	11 µl
2x Hybridization Mix	110 µl
DMSO	15.4 µl
Nuclease-free water	19.9 µl
Total Volume	160 µl

The master mix was thoroughly homogenised and 160µl of the master mix was added to 60µl (5.2 µg) of the fragmented and biotin-labelled ss-cDNA libraries. Reactions were mixed and brief centrifuged. Libraries were incubated in a thermocycler using the recommended program (99ºC for 5 minutes and 45ºC for 5 minutes) with heated lid (@ 99ºC). After incubation, 200µl of the each library was injected and hybridized to HT 2.0 expression array using GeneChipR Hybridization Oven 645 (Figure 20). A pipette tip was inserted into upper right septum for venting while the samples were loaded from the bottom left septum on the array. After loading the libraries, both septa were covered with tough spot stickers to minimize evaporation and to prevent leaks. The arrays were loaded on the hybridization trays in a rotating incubator (rotation speed 60 RPM) for 16 hours at 45ºC.

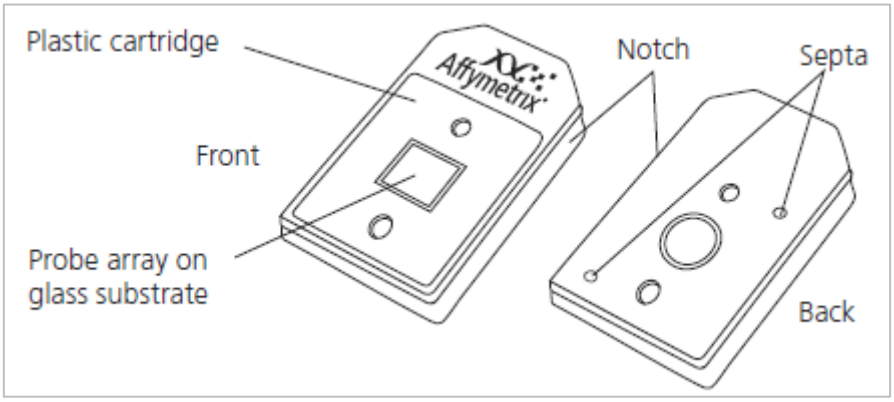


Figure 20: Components of the HTA 2.0 array chip.

The square in the centre indicates the probe arrays where libraries are loaded.

After 16 hours of incubation, HTA 2.0 arrays were removed and the hybridization cocktail was aspirated through the septum using a pipette. 250µl of wash solution-A was loaded into each array and the arrays were placed into the GeneChipR Fluidics Station 450. Following instructions on the LCD of the fluidics station, 1.5ml of Eppendoff tube containing a master mix (600µl Stain Cocktail 1, 600µl Stain Cocktail 2 and 800µl of array holding buffer) was placed on each holder. Following on, arrays were washed and stained in the fluidics station for 90 minutes. This Human Transcriptome 2.0 gene expression profiling was performed under the supervision of Dr. Azim Mohamedali.

2.2.15.10 Expression microarray data analysis

Data analysis was performed using Affymetrix expression console and Affymetrix transcriptome console software (Figure 21). The HT 2.0 arrays were scanned using a GeneChipR 3000 Scanner (7G upgrade) to capture images (DAT files). The raw images from the DAT file generated from Affymetrix array probes were converted to CEL files. Using the Affymetrix expression Console software (build 1.3.1.187), probe cell intensity files (*.CEL) were converted to probe level summarization files (*.CHP) for all 12 samples ($CD34^+$ samples, n=4; Erythroid $CD71^+CD235^+$ samples, n=4). Following on, initial data analysis (including annotation, data normalization and data QCs) was performed using Affymetrix expression console with default settings.

All the CHP files were imported into the Affymetrix transcriptome analysis console (TAC) 2.0 (build 2.0.0.9) using Gene Level and Exon Level differential analysis option. Within the TAC, the genome assembly hg19 was used with the gene-level and exon-level library annotation file supplied by Affymetrix.

TAC was further used to perform both "gene" level (one-way ANOVA analysis) and "splicing" level (1-way ANOVA analysis) differential expression analysis between the *SF3B1* mutant vs *SF3B1* wildtype patient samples ($CD34^+$ cells and $CD71^+CD235^+$ cells). The data was filtered further using a FDR <0.1 (unadjusted P value) and a fold-change cut-off of ≥±2 (or Splicing Index ±2) to create a list of significantly up or down regulated genes and differentially splicing gene transcripts.

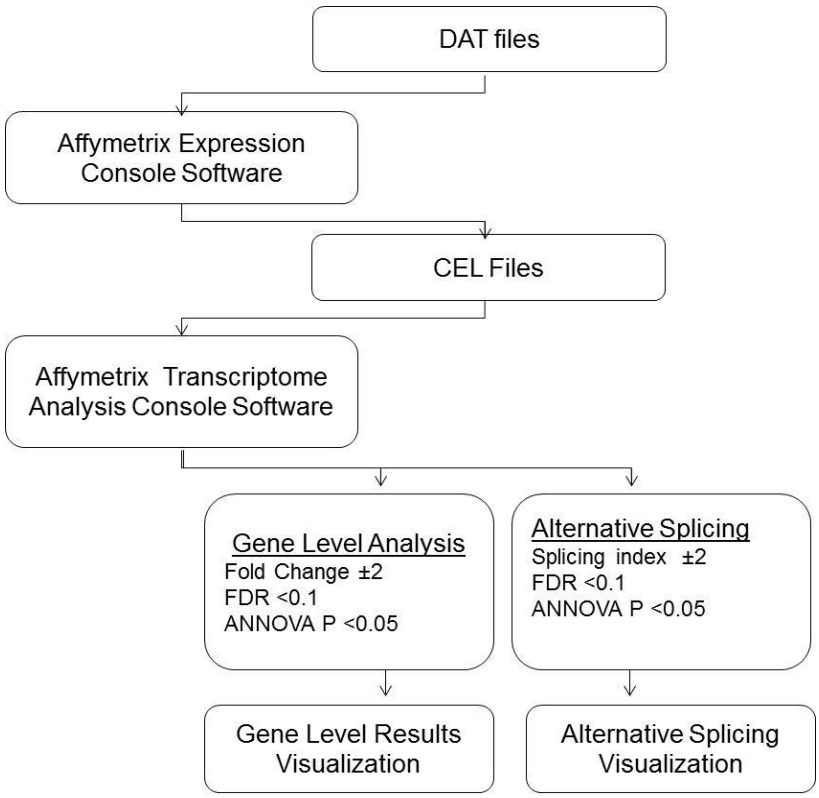


Figure 21: Data processing pipeline for Affymetrix HTA2 array data analysis.
DAT files were converted to CHP files using Affymetrix expression console. Normalized and annotated CEL files were then imported into the Affymetrix transcriptome analysis console (TAC). Gene level and alternative splicing analysis was performed using TAC. Filtration threshold were applied and then data was visualized manually.

Chapter 3

SF3B1 is a molecular marker for MDS patients with Ring Sideroblasts

3.1 Whole- exome sequencing of Eleven MDS-RS patients

RARS is a low-risk MDS subtype that is characterized by isolated anaemia, erythroid dysplasia and less than 5% myeloblasts. Furthermore, 15% or more ring sideroblasts are characteristically observed in the differentiating bone marrow erythroid precursors. However, ringed sideroblasts are also characteristic of other inherited sideroblastic anaemias (Cazzola and Invernizzi, 2011). Aberrations of various genes involved in haem biosynthesis, iron-sulfur (Fe-S) cluster biogenesis, or Fe-S cluster transport, and mitochondrial metabolism have been implicated in these inherited disorders (Bergmann et al., 2010; Cotter et al., 1999; Maguire et al., 2001; Smith et al., 1995). Interestingly, none of these candidate genes have been found to be mutated in RARS, although various authors have proposed the role of mitochondrial dysfunction. Furthermore, DNA methylation and histone modification abnormalities are frequently observed in the bone marrow cells in these patients that can itself lead to altered gene expression and genomic instability (Issa, 2010), therefore suggesting that RARS disease occurs as a result of multiple and cumulative alterations to the hematopoietic stem cells. Additionally, stem cells in RARS patients have also been shown to be clonal, therefore providing the evidence for the clonal nature of this disorder. Therefore, the initiating events (genetic or/and epigenetic) responsible for the disease not only provides proliferative advantage to the abnormal clone, but also leads to ineffective erythropoiesis and dysfunctional erythroid iron metabolism causing accumulation of the metal in the mitochondria. Since the abnormal phenotype i.e. bone marrow dysplasia observed in RARS is in the erythoid lineage only, suggesting a genetic homogeneity in this disease group and hence presents a strong likelihood of having a common genetic cause although this has not however been determined yet. Therefore, RARS provides an ideal disease sub-group for further mutation screening of the protein coding part (whole-exome) of the genome.

Sequencing of complete genome of an individual with AML using ‘second-generation’ sequencing (SGS) technology represented a landmark in genetic screening of the leukaemia genome. This SGS technology along with the new DNA-enrichment technologies, such as array/bead hybridization-based region capture, has enabled highly efficient targeted sequencing of regions of interest, hence increasing numbers of samples that can be processed simultaneously. This focussed way of sequencing

the smaller fraction of the genome such as protein coding exons (the 'exome') enables us to sequence the targeted region at a higher depth (reads per base) with no extra cost. The haploid human exome roughly constitutes ≈1% of the human genome spanning across ≈180,000 exons. One of the challenges for focusing on the exonic regions has been how to define the set of targets that constitute the exomic regions within the whole-genome. In addition, there has been considerable uncertainty regarding which genomic sequences of the human genome are truly protein coding regions. Initial commercial whole-exome capture kits used the CCDS (Consensus Coding Sequence Project) definition of the exome, which is subset of genes determined to be coding with high confidence. However, recently available commercial kits now target all of the genes present in the Refseq database (Pruitt et al., 2007) and an increasingly large number of hypothetical proteins. In at least one example, identification of a novel disease gene *MLL2* shown to be the causal marker in Kabuki syndrome (Ng et al., 2010) would have been missed, if the CCDS definition was not expanded to the Refseq database (Pruitt et al., 2007). Reflecting this, more recent whole-exome kits now target up to ~62Mb region, rather ~30Mb.

With this approach, genomic DNA is fragmented and then hybridized to capture RNA probes that are specific to the exonic regions of the genome. Following on, capture genomic region is isolated and purified by magnetic beads or other purification strategies. There are other methods which are used for targeted enrichment and these include amplicon/targeted PCR or capture by circularization. Sequencing this region substantially reduces the amount of sequencing necessary to ascertain protein coding variants with higher confidence. In addition, by focussing only on the exonic regions of the genome we avoid sequencing the majority of the repetitive regions within the genome, which are difficult to map, assemble and call variants from. WES studies have demonstrated that this approach is a cost-effective means of precisely denoting functional variations in a large cohort of samples (Meyerson et al., 2010; Otto et al., 2010). Furthermore, previous studies have demonstrated that some genomic alterations in MDS are present at low frequencies (<20% mutant allele burden) in clinical samples, hence making it difficult to detect these gene mutations. This is largely due to the substantial admixture of MDS dysplastic haematopoietic cells within the different types of myeloid cells and normal lymphoid cell populations (such as T-

cells and B-cells) in the bone marrow compartment, thus making it difficult to molecularly characterize and model the disease. Sequencing only the exonic regions at higher depth (reads per base) will also help us to detect mutation present at lower frequencies (<20% mutant allele burden). Therefore, applying SGS technology to perform WES on MDS-RS will allow a comprehensive and unbiased investigation of the complete protein-coding regions, and will therefore help to identify novel disease specific acquired and/or inherited mutations. This will enable us to unravel the potential molecular mechanism of the phenotypic abnormalities which in-turn will help to elucidate the pathophysiology, refine the prognostic scoring systems, and more importantly provide novel therapeutic targets for this subgroup of MDS.

Since MDS in-general is specific to myeloid lineage of the haematopoiesis, therefore DNA isolated from patient's $CD34^+$ cells is the ideal source of material for whole-exome screening. Eleven MDS-RS patients all of whom had >25% ringed sideroblasts patients, comprising 7 RARS, 2 RARS-T and 2 RCMD-RS (Table 24) were selected for the initial whole-exome sequencing. One patient with congenital sideroblastic anaemia was also included in the WES analysis. $CD34^+$ cells were isolated from bone marrow cells of these patients using MACS isolation protocol (see methods section 2.2.8 for more details). Previous studies have reported that a purity of ≥94% can be achieved with this isolation method (Bonanno et al., 2009; Giarratana et al., 2005).

Analysis of constitutional DNA in patients with haematological malignancies is essential to delineate the origin of a particular genetic lesion. This will enable us to distinguish the somatic alterations from germline mutations or polymorphisms. Furthermore, possible sources of constitutional DNA which can be used are skin biopsies, cultured skin fibroblasts, $CD3^+$ T-cells bone marrow or peripheral blood cells and buccal epithelial cells. However, DNA obtained from patient's buccal epithelial cells can be contaminated with leukemic DNA, hence suggesting that this DNA source is not an ideal source of constitutional DNA (Endler et al., 1999; Zebisch and Sill, 2008). Therefore, paired skin biopsy or bone marrow $CD3^+$ T-cells from patients were used as constitutional source of DNA in this study. $CD3^+$ T cells were isolated from patient's bone marrow cells by MACS and following the same protocol which was used for $CD34^+$ isolation (see methods section 2.2.8 for more details).

WES was performed on $CD34^+$ cells (primary bone marrow sample) as described in methods section 2.2.4. Paired constitutional DNA from skin (6 patients) or $CD3^+$ T-cells (4 patients) was similarly sequenced in parallel. An average of 8-13 gigabase of sequence data was generated per patient exome and processed using the Burrows-Wheelers aligner (BWA)/GATK pipeline (Broad Institute 'Best practise' pipelines). Additional software tools were also used for downstream data analysis as detailed in methods section 2.2.4.6 (See methods section for full details). The mean base coverage across the targeted region ranged from 162 reads to 325 reads, with >98% of the exome base coverage of >10X in all cases (Figure 22). Functional variants not found in dbSNP140, esp5400 (<0.01%) and 6500 (<0.01%) or 1000 genomes (<0.01%) databases were subsequently collated.

Table 24: Total variant (SNVs and Indels) with ≥3 tumour mutant reads observed for each individual patient were subjected to our filtration criteria as discussed in methods chapter. Candidate filtered somatic variants generated were selectively subjected to confirmation using Sanger sequencing. For patient with no constitutional source of DNA, only known myeloid related gene mutations were passed for further confirmation. No somatic variants were observed for one patient with congenital sideroblastic anaemia. Only candidate missense, nonsense, splice-site and indel variants were subjected to confirmation. †Congential Sideroblastic anaemia, *Patient sample with no paired constitutional DNA source.

Patients	SNVs						Indels			
	Unfiltered Data		Filtered Data				Unfiltered Data		Filtered Data	
	Total SNVs	Exonic SNVs	Synonymous	Missense	Nonsense	Splice-site	Total Indels	Exonic Indels	In-frame	Frameshit
MDS1	123,914	21,891	23	11	0	0	13,474	421	4	0
MDS2	115,277	21,044	1	14	0	1	25,571	867	0	0
MDS3	103,812	23,415	6	5	0	0	14,807	686	1	1
MDS4	93,110	22,115	2	4	0	0	12,321	572	0	0
MDS5†	93,435	22,877	0	0	0	0	11,897	661	0	0
MDS6	131,607	22,777	0	10	1	0	16,481	430	0	0
MDS7	155,482	23,769	3	11	2	0	18,678	491	0	3
MDS8	146,862	23,026	4	12	1	0	18,186	442	2	1
MDS9	121,353	22,372	0	8	0	0	15,207	409	1	1
MDS10	128,628	22,809	2	4	0	0	15,913	440	1	2
MDS11*	132,578	25,462	N/A	N/A	N/A	N/A	16,212	463	N/A	N/A
MDS12	142,734	22,794	0	8	0	0	17,770	456	0	1

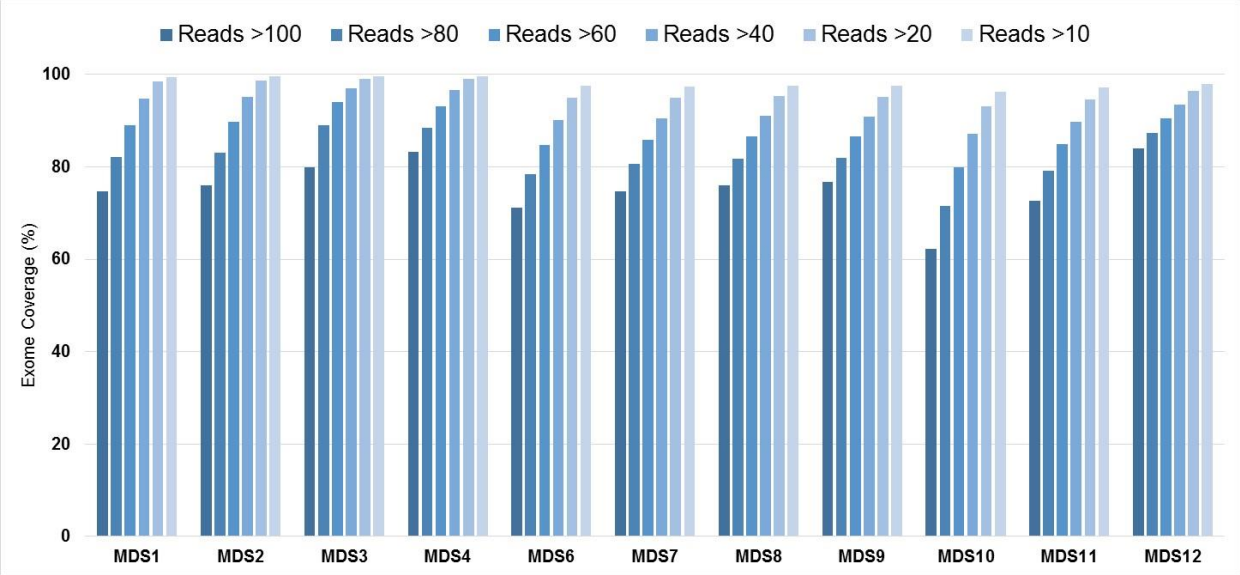


Figure 22: Data coverage of the WES data.

The percentage of targeted bases that were covered by sequencing reads. More than 65% of the exome was covered with at-least 100 sequencing reads and 98% of the exome was covered with ≥20 sequencing reads.

Due to the late onset of the phenotypic as well as clinical presentation of the patients with MDS (median age of diagnosis 70-75 years), we hypothesised that any genetic aberration which leads to the disease must be a somatic lesion acquired during the lifetime of the patient. In addition, none of the inherited mutations present in congenital sideroblastic anaemia patients have reported in patients with acquired sideroblastic anaemia. Therefore, we decided to concentrate only on the acquired mutations present in the patients $CD34^+$ cells. Following on, somatic mutations were subsequently passed when somatic p-value (VarScan, Fisher's exact t-test) was <0.01, had ≥3 reads supporting mutation, were present in paired normal tissue at less than 20% of the paired tumour tissue skin (or <50% for T-cells) and had a mutant allele burden (MAB) of >5%. The reason for selecting this mutation ratio between normal/tumour tissues for initial analysis was to avoid false negative results which could be due to the fact that skin biopsies and $CD3^+$ T-cells from patients might have background tumour cell contamination.

Following on from initial whole-exome data analysis, acquired candidate mutations were identified for 10 patients with paired constitutional DNA samples. These mutations were selectively validated by Sanger sequencing (Table 25). For 1 patient

where paired constitutional DNA was not available, only known myeloid-related genes mutations were included for further analysis. The total number of confirmed somatic mutations were 54, ranging from 2 to 10 per patient. This mutation rate was comparable to that in AML (7.3 – 13) (Ley et al., 2010) and chronic lymphocytic leukaemia (≈12) (Puente et al., 2011), however, it is lower than multiple myeloma (≈33)(Chapman et al., 2011) and solid tumours (32-500) (Lee et al., 2010b; Shah et al., 2009). Of the mutations detected in our patient cohort, 4 were frameshift insertion-deletions (indels) and 50 were non-synonymous; 49 were found in coding sequences and one in intron splice-junction (within 3 bp from the exon junction). The mutation spectrum observed in our cohort showed a predominance of transitions, especially C→T and G→A. This mutational spectrum is similar overall to those observed in haematological, colorectal, pancreatic, and brain cancers (Greenman et al., 2007; Jones et al., 2008; Papaemmanuil et al., 2011; Puente et al., 2011; Yoshida et al., 2011b).

Table 25: Detailed information for somatic mutations identified by whole-exome sequencing in 11/12 patients.

Complete list of somatic mutation detected in patients with a paired constitutional DNA are shown in table. However, patient where no constitutional DNA was available (n=1), only known myeloid specific somatic mutations are shown in table. Gene names in bold black font and with (*) represent the known myeloid related gene mutations. Gene names in bold red font represent the gene commonly mutated in these patients.

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
MDS1	***SF3B1***	ENSP 00000335321	chr2: 198266834	T	C	Lys700Glu	43%	Skin	yes	0	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS1	*HIST1H2AC*	ENSP 00000321389	chr6: 26124820	G	T	Lys120Asn	31%	Skin	no	0.01	* Damaging	Benign	Core component of nucleosome
MDS1	*FBLIM1*	ENSP 00000364921	chr1: 16091611	G	A	Ala45Thr	27%	Skin	no	0.35	Tolerated	Benign	Implicated in cell shape modulation (spreading) and motility
MDS1	*RETNLB*	ENSP 00000295755	chr3: 108475355	C	T	Gly70Arg	28%	Skin	no	0.03	Damaging	Probably damaging	Probable hormone
MDS1	***JAK2****	ENSP 00000371067	chr9: 5073770	G	T	Val617Phe	36%	Skin	yes	0	Damaging	Probably damaging	Non-receptor tyrosine kinase involved in various processes such as cell growth, development, differentiation or histone modifications
MDS1	*TH*	ENSP 00000370571	chr11: 2189733	C	T	Ala159Thr	15%	Skin	no	0.02	Damaging	Benign	Plays an important role in the physiology of adrenergic neurons
MDS2	***SF3B1***	ENSP 00000335321	chr2: 198267371	G	C	His662Gln	42%	CD3+ T-cells	yes	0.01	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
													snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS2	*G3BP1*	ENSP 00000377681	chr5: 151175137	G	C	SS Exon 6 G>C	40%	CD3+ T-cells	no	N/A	N/A	N/A	DNA-unwinding enzyme
MDS2	***DNMT3A****	ENSP 00000370122	chr2: 25467133	C	G	Trp581Ser	41%	CD3+ T-cells	no	1	Tolerated	Probably damaging	Required for genome-wide de novo methylation
MDS2	*TLL2*	ENSP 00000350630	chr10: 98145885	A	G	Val647Ala	7%	CD3+ T-cells	no	0.15	Tolerated	Probably damaging	Protease which specifically processes pro-lysyl oxidase. Required for the embryonic development
MDS2	*NMNAT2*	ENSP 00000294868	chr1: 183230405	C	T	Val224Met	41%	CD3+ T-cells	no	0.02	Damaging	Probably damaging	Catalyses the formation of NAD(+) from nicotinamide mononucleotide (NMN) and ATP
MDS2	*NAGA*	ENSP 00000379680	chr22: 42463872	T	A	Tyr74Phe	35%	CD3+ T-cells	no	0.05	Damaging	Possibly damaging	Required for the breakdown of glycolipids
MDS2	*ABCB11*	ENSP 00000263817	chr2: 169833167	C	T	Gly410Ser	38%	CD3+ T-cells	no	0	Damaging	Probably damaging	Involved in the ATP-dependent secretion of bile salts into the canaliculus of hepatocytes
MDS2	*NR0B2*	ENSP 00000254227	chr1: 27238545	C	T	Gly189Arg	29%	CD3+ T-cells	no	0.03	Damaging	Benign	Acts as a transcriptional regulator
MDS2	*SPATA13*	ENSP 00000371576	chr13: 24797870	G	A	Arg268Gln	41%	CD3+ T-cells	no	0.73	Tolerated	N/A	Acts as guanine nucleotide exchange factor (GEF) for RHOA, RAC1 and CDC42 GTPases
MDS2	*CCDC17*	ENSP 00000389415	chr1: 46088498	A	G	Leu213Pro	17%	CD3+ T-cells	no	0.29	Tolerated	Benign	Unknown
MDS3	***SF3B1***	ENSP 00000335321	chr2: 198267371	G	T	His662Gln	28%	Skin	yes	0.01	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
MDS3	*ME1*	ENSP 00000358719	chr6: 84108234	A	G	Try72His	31%	Skin	no	0.32	Tolerated	Probably damaging	Unknown
MDS3	*CAMTA1*	ENSP 00000306522	chr1: 7731086	G	T	Cys923Phe	33%	Skin	no	0.02	Damaging	Possibly damaging	Transcriptional activator. May act as a tumour suppressor
MDS3	*PTPDC1*	ENSP 00000288976	chr9: 96860488	A	T	Pro357Leu	21%	Skin	no	0.1	Tolerated	Benign	May play roles in cilia formation and/or maintenance
MDS4	***SF3B1***	ENSP 00000335321	chr2: 198266834	T	C	Lys700Glu	40%	Skin	yes	0	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS4	*MYH14*	**ENSP 00000262269**	chr19: 50728875	C	G	Pro251Ala	44%	Skin	no	0	*Damaging	Probably damaging	Cellular myosin that appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping
MDS6	*FGD1*	ENS P00000364277	chrX: 54492200	G	A	Arg476Trp	23	Skin	no	0	Damaging	Probably damaging	Activates CDC42, a member of the Ras-like family of Rho- and Rac proteins. Plays a role in regulating the actin cytoskeleton and cell shape.
MDS6	*TESC*	ENSP 00000432716	chr12: 117479796	G	A	Pro148Ser	25	Skin	no	0.19	Tolerated	Probably damaging	Functions as an integral cofactor in cell pH regulation, cell maturation, cellular transport and cell surface stability.
MDS6	***SF3B1***	ENSP 00000335321	chr2: 198267491	C	G	Glu622Asp	40	Skin	yes	0.05	Damaging	Possibly damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS6	***DNMT3A***	ENSP 00000264709	chr2: 25463286	C	T	Arg736His	8	CD3+ T-cells	yes	0.39	Tolerated	Probably damaging	Required for genome-wide de novo methylation
MDS7	*ATP1A4*	ENSP 00000357060	chr1: 160143980	C	A	His691Asn	24	CD3+ T-cells	no	0	* Damaging	Possibly damaging	Part of ion pump responsible for maintaining sodium and potassium

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
													electrochemical gradients across the plasma membrane
MDS7	*MCM3AP*	ENSP00000380820	chr21: 47666752	G	A	Gln1447X	23	CD3+ T-cells	no	N/A	Tolerated	N/A	Involved in the nuclear localization pathway of MCM3
MDS7	***SF3B1***	ENSP00000335321	chr2: 198267371	G	C	His662Gln	45	CD3+ T-cells	yes	0.01	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS7	*GPR112*	ENSP00000377699	chrX: 135455180	T	C	Met2578Thr	25	CD3+ T-cells	no	0.34	Tolerated	Benign	Orphan receptor
MDS8	*PKHD1*	ENSP00000360158	chr6: 51947297	C	A	Leu58Phe	37	Skin	no	0.27	Tolerated	Probably damaging	May be required for correct bipolar cell division through the regulation of centrosome duplication and mitotic spindle assembly
MDS8	*MC2R*	ENSP00000333821	chr18: 13885474	G	A	Ala15Val	63	Skin	no	0.05	Damaging	Benign	Member of the five-member G-protein associated melanocortin receptor family
MDS8	*PCNT*	ENSP00000352572	chr21: 47836271	A	G	Ile2147Val	40	Skin	no	1	Tolerated	Benign	Integral component of the filamentous matrix of the centrosome involved in the initial establishment of organized microtubule arrays in both mitosis and meiosis
MDS8	***SF3B1***	ENSP00000335321	chr2: 198267371	G	C	His662Gln	37	Skin	yes	0.01	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS8	*ATG9B*	ENSP00000475005	chr7: 150721444	C	T	Gly23Arg	44	Skin	no	N/A	N/A	Benign	Plays a role in the regulation of autophagy
MDS8	*FITM2*	ENSP00000380037	chr20: 42935567	C	T	Ala163Thr	48	Skin	yes	0.8	Tolerated	Benign	Plays an important role in lipid droplet accumulation. Plays a role in the regulation of cell morphology and cytoskeletal organization
MDS8	*AFF3*	ENSP00000386834	chr2: 100209883	G	T	Pro747His	19	Skin	no	0.09	Tolerated	Brobably damaging	Putative transcription activator that may function in lymphoid development and

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
													oncogenesis. Binds, in vitro, to double-stranded DNA
MDS 8	*TET2*[*]	ENSP 00000442788	chr4:106164 862	CT	-	1244_1244d el	45	Skin	no	0.03	Damaging	N/A	Dioxygenase that catalyses the conversion of the modified genomic base 5mC into 5hmC and plays a key role in active DNA demethylation
MDS8	*CEP110*	ENSP 00000362972	chr9: 123860787	A	G	Ser249Gly	25	Skin	no	0.33	Tolerated	Probably damaging	Required for centrosome duplication at different stages of procentriole formation.
MDS9	*AGXT2L1*	ENSP 00000296486	chr4: 109680914	G	A	Thr103Ile	37	Skin	no	0.08	Tolerated	benign	Catalyses the pyridoxal-phosphate-dependent breakdown of phosphoethanolamine
MDS9	*IKBKE*	ENSP 00000356087	chr1: 206658341	A	G	Ser394Gly	29	Skin	no	0.41	Tolerated	Benign	Phosphorylates inhibitor of NF-kappa-B
MDS9	*GAB4*	ENSP 00000383431	chr22: 17472858	C	T	Arg128His	40	Skin	yes	0	Damaging	Probably damaging	Unknown
MDS9	*OC90*	ENSP 00000254627	chr8: 133036862	C	T	Ala434Thr	37	Skin	no	0.77	Tolerated	Benign	Unknown
MDS9	***SF3B1***	ENSP 00000335321	chr2: 198266494	T	C	Asp781Gly	45	Skin	yes	0	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS1 0	*ANXA7*	ENSP 00000362012	chr10: 75143371	C	-	Thr259fs	26	CD3+ T-cells	no	0	Damaging	N/A	Calcium/phospholipid-binding protein which promotes membrane fusion and is involved in exocytosis
MDS1 0	***SF3B1***	ENSP 00000335321	chr2:198266 834	T	C	Lys700Glu	40	CD3+ T-cells	yes	0	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS1 0	***DNMT3A***[*]	ENSP 00000264709	chr2:254572 42	C	T	Arg882His	50	CD3+ T-cells	yes	0.03	Damaging	Possibly damaging	Required for genome-wide de novo methylation
MDS1 0	*SLC7A2*	ENSP 00000419140	chr8:174178 99	C	T	Ser494Leu	39	CD3+ T-cells	no	0.01	Damaging	Benign	Involved in the transport of the cationic amino acids

Patient ID	Gene	Ensembl Protein ID	Chromosome Location	Reference Base	Mutant Base	Codon Change	MAB (%)	Constitutional DNA Source	COSMIC	SIFT Score	SIFT Prediction	PolyPhen	Gene/Protein function
MDS11	***SF3B1***	ENSP00000335321	chr2: 198267491	C	A	Glu622Asp	60	N/A	yes	0.05	Damaging	Possibly damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS11	***DNMT3A****	ENSP00000264709	chr2: 25457242	C	T	Arg882His	75	N/A	yes	0.03	Damaging	Possibly damaging	Required for genome-wide de novo methylation
MDS12	*KIAA1109*	ENSP00000264501	chr4: 123109050	G	A	Gly 210Arg	40	CD3+ T-cells	no	0.01	Damaging	Probably damaging	Unknown
MDS12	***SF3B1***	ENSP00000335321	chr2: 198266834	T	C	Lys 700Glu	43	CD3+ T-cells	yes	0	Damaging	Probably damaging	Subunit of the splicing factor SF3B required for 'A' complex assembly formed by the stable binding of U2 snRNP to the branch-point sequence (BPS) in pre-mRNA
MDS12	***TET2****	ENSP00000442788	chr4: 106157287	A C	-	730_730del	42	CD3+ T-cells	no	0.02	Damaging	N/A	Dioxygenase that catalyses the conversion of the modified genomic base 5mC into 5hmC and plays a key role in active DNA demethylation
MDS12	*PPM1D*	ENSP00000306682	chr17: 58740624	A	-	Gln510fs	42	CD3+ T-cells	no	0.04	Damaging	N/A	Required for the relief of p53-dependent checkpoint mediated cell cycle arrest
MDS12	***DNMT3A****	ENSP00000264709	chr2: 25466800	G	A	Arg635 Trp	50	CD3+ T-cells	yes	0	Damaging	Probably damaging	Required for genome-wide de novo methylation

3.1.1 Whole-exome sequencing reveals *SF3B1* mutations in MDS-RS

The molecular pathogenesis and the cause of the accumulation of mitochondrial iron that gives rise to ringed sideroblasts in MDS-RS patients is unknown. We evaluated whether WES could be applied to identify directly the causative gene underlying the ring sideroblastic phenotype. Even in this scenario for 'whole exome data set', the key challenge that arises immediately is that the large number of mutations present by chance in any single 'diseased genome' makes it difficult to identify which variant is causal, even when only considering somatic functional variants. Because of the genetically homogenous phenotype observed in the MDS-RS patients, we hypothesized that any genetic defect leading to ring sideroblatic phenotype would involve a common gene(s) or a pathway.

Among the genes found to be recurrently mutated in our cohort of patients, splicing factor 3b subunit 1 (*SF3B1*), a component of the major and minor spliceosomes, initially stood out due to its high mutation frequency. *SF3B1* mutation was present in the 11/12 patients sequenced in this study (Table 25). Four patients carried Lys700Glu mutation, three patients His662Gln mutation, two patients Glu622Asp mutation and one patient had Asp781Gly mutation in the *SF3B1* gene. On the basis of the proportion of reads reporting the mutant allele burden (MAB), all the mutations were heterozygous and present as a dominant clones. Validation of these mutations in available paired skin biopsy by Sanger sequencing allowed me to confirm acquired nature of these mutations (Figure 23). No acquired mutation was detected in one patient (MDS5) with congenial sideroblastic anaemia, however, a constitutional *ALAS2* (R425C) gene mutation which has been previously (Camaschella, 2009; Cotter et al., 1999; Kaneko et al., 2013) linked with sideroblast phenotype was present in this patient.

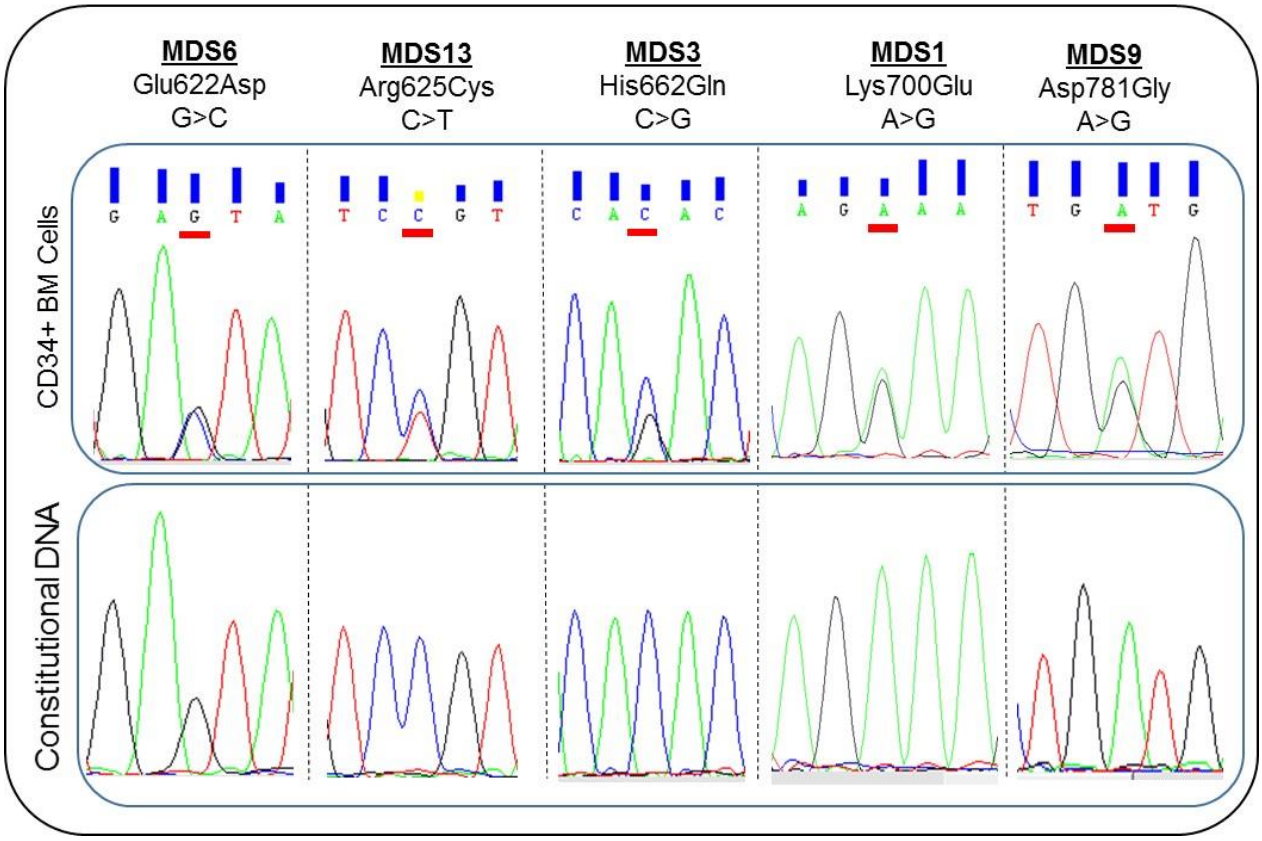


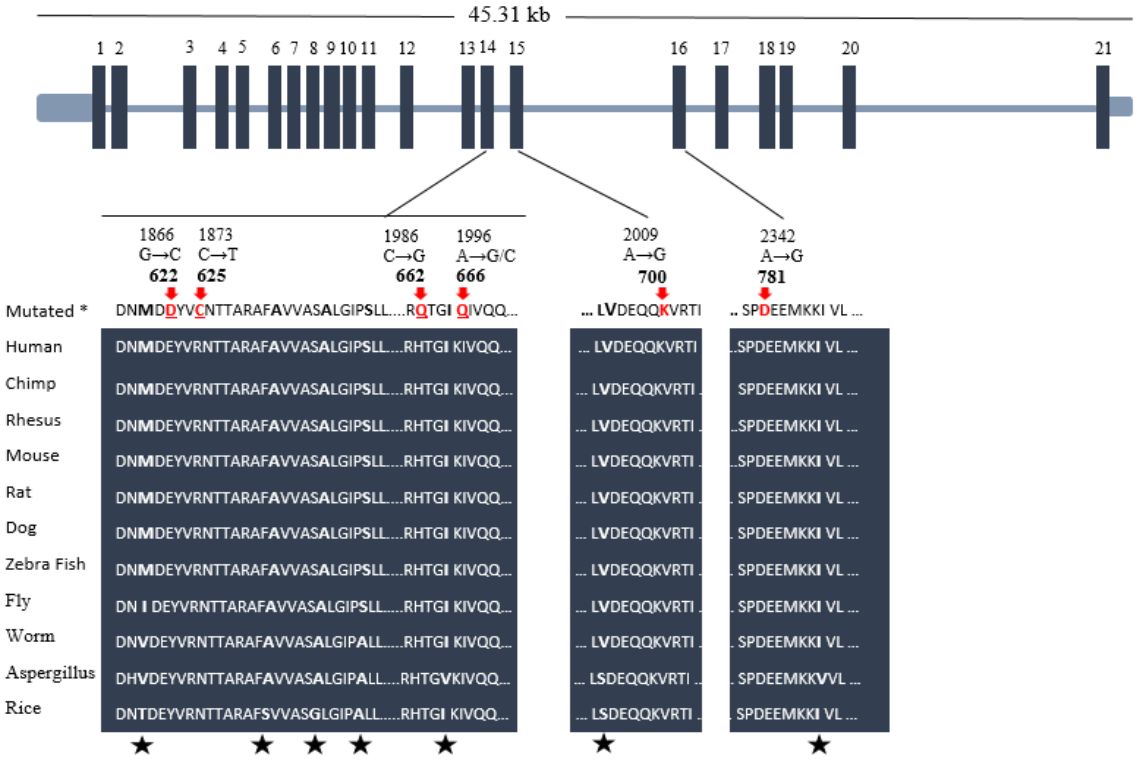
Figure 23: Somatic *SF3B1* mutations in MDS patients with ring sideroblasts.

Sanger sequence electropherogram (top panel) from 5 patients demonstrates the most common mutation identified in this study. The absence of these mutation in a sequence electropherogram from matched constitutional DNA (bottom panel) obtained from skin biopsy confirms the somatic nature of these mutations.

To characterize the spectrum and frequency of *SF3B1* mutations in greater detail, targeted mutational sequencing of the *SF3B1* gene was performed in 18 additional MDS-RS patients (RARS, n=7; RARS-T, n=1; RCMD-RS, n=9 and MDS/AML, n=1). All the patients had >15% ring sideroblast at the time of sample collection. *SF3B1* mutations were identified in 86% (6/7) RARS, 100% (1/1) RARS-T, 44% (4/9) RCMD-RS and 100% (1/1) MDS/AML with a preceding RARS (Data was presented as a poster at 53rd American Society of Haematology Conference, 2011, San Diego, California). At the same time, similar data was published by two other groups showing the presence of *SF3B1* mutations in MDS-RS patients (Papaemmanuil et al., 2011; Yoshida et al., 2011b).

3.1.2 *SF3B1* mutations- Type, allele burden and sites

The distribution of observed mutations across the *SF3B1* gene was striking. All *SF3B1* mutations were non-synonymous amino-acid substitutions. MAB ranged from 28% to 60% with an average mutant allele burden of 45% (n=29), indicative of a heterozygous state. No frameshift, indels, splice-site mutations, or nonsense substitutions were observed. All the mutations clustered in exons 14 to 16 of the gene, and one variant in particular, Lys700Glu, accounted for 12 of the 23 variants observed (Figure 24). Several other mutated amino acid residues were also observed, these include His662 (n=5), Glu622 (n=2), Arg625 (n=1), Lys666 (n=2), and Asp781 (n=1). All *SF3B1* mutations clustered in the protein c-terminal HEAT motifs implicated in snRNP stabilization within the U2 snRNP complex of the major spliceosome.



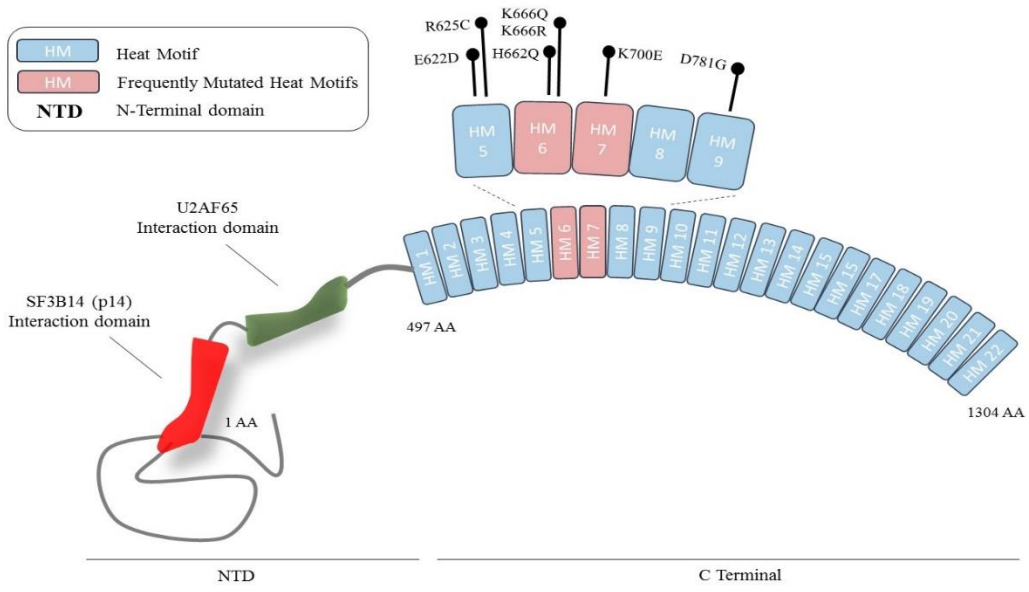


Figure 24: Somatic SF3B1 mutations in MDS patients with ring sideroblasts.

(A) *SF3B1* gene with all known exons and protein sequence alignments of the *SF3B1* C-terminal domain around the mutated residues (Red arrows) in evolutionarily diverse species. (B) Distribution of somatic mutations across the *SF3B1* protein. Top panel represents the mutated HEAT motifs and the mutated codons of the protein. NTD, N-terminal domain.

3.1.3 *In-silico* analysis of *SF3B1* mutations

To investigate the impact of mutations, an *in-silico* 3-D model of the C-terminal part of the *SF3B1* protein (Figure 25) was constructed using SWISS-MODEL (Biasini et al., 2014). In this model, most of the aberrations clustered in the fifth, sixth, seventh, eighth and ninth HEAT motifs on the inner surface of the protein and may define a binding interface, consistent with the suggestion that the helical motifs form the outer shell of the U2 complex. Interestingly, the sixth motif which contains the frequent Lys700Glu mutations is at the hinge (Figure 24 and Figure 25) of the S-shape shell like structure. Similarly, Asp781Gly also affects the external loop at the hinge. As this part of the protein is highly conserved among various species, therefore it is highly likely that these mutation are deleterious.

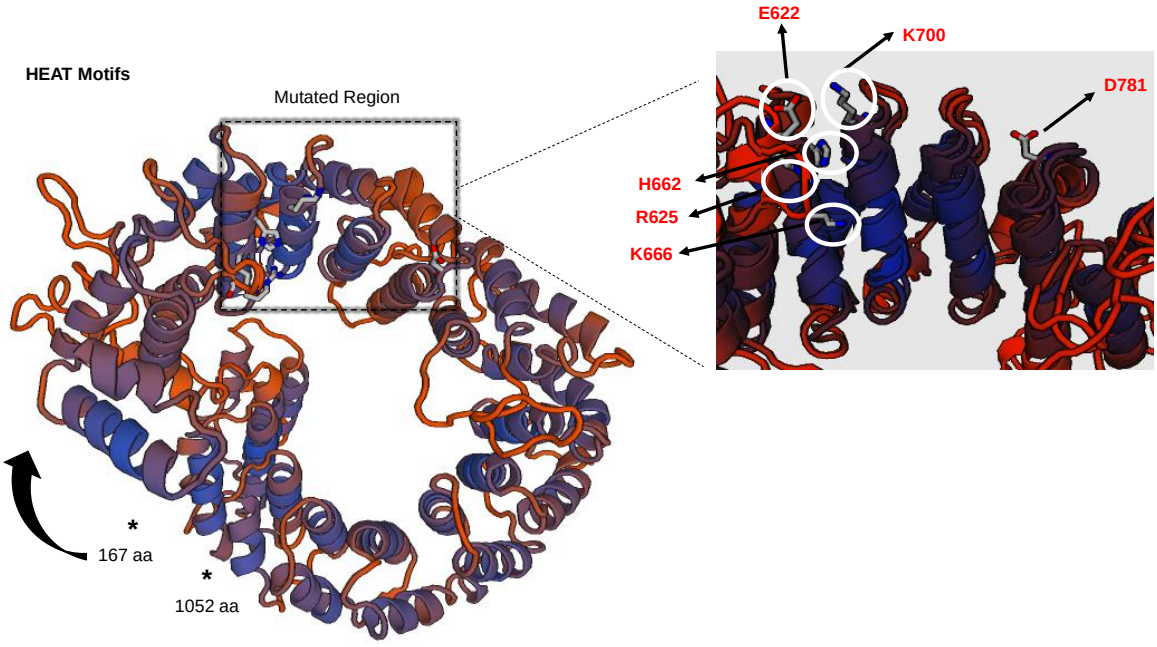


Figure 25: 3D structure of C-terminal domain of the Sf3b1 protein (left) and detailed view of the mutated codons (right).

Each loop represent a heat domain present within the N-terminal of the Sf3b1. Mutated codons are represented within the white circles (Right) by the small structures (Red and Blue).

3.1.3.1 *SF3B1* – Transversion mutations

The 1866 G→C mutation leads to the replacement of Glutamic acid (Molar mass, Mm-147.12926 g/mol) with Aspartic acid (Mm-133.10268 g/mol) at codon 622 while 1986 C→G mutation results in Histidine (Mm-155.15456 g/mol) to Glutamine (Mm-155.15456 g/mol) replacement at codon 662. The wild-type and mutant amino acids in both the mutations differ in size (Figure 26). Mutant residues appear to be smaller in mass which might lead to loss of interactions with other interacting proteins.

The 1996 A→G mutation results in Lysine (Mm- 146.18756 g/mol) to Arginine (Mm-174.20096 g/mol) substitution at codon 666. Both the wildtype and mutant residues share the same backbone (Figure 26), however the side chain is different. The mutant residue is larger than the wild-type residue which might lead to irregularities in the protein structure. However, 1996 A→C mutation results in the Lysine (Mm- 146.18756 g/mol) to Glutamine (Mm-146.1445 g/mol) substitution at codon 666. In this case, there is a difference in charge with the wild-type residue being positively charged while the mutant residue is neutral. The loss of charge can cause changes in the interactions with other molecules or residues. The wild-type and mutant amino acids also differ in size. The mutant residue is smaller which could lead to loss of interaction with other interacting proteins.

3.1.3.2 *SF3B1* - Transition mutations

The 2009 A→G mutation leads to Lysine to Glutamic acid substitution at codon 700 and both residues (wildtype and mutant) have the same backbone (Figure 26). The mutant residue is small and negatively charged while as the wildtype residue is larger with a positive charge. This change in the charge and size can cause repulsion with other residues in the protein or ligands.

The 2342 A→G mutation changes Aspartic acid to Glycine at codon 781. The wild-type residue is negatively charged while as the mutant residue is neutral. However, 1873 C→T mutation results in the substitution of Arginine into Cysteine at codon 625 with a positive charge of the wildtype residue and a neutral charge for mutant residue. The backbone of the wildtype and mutant in both cases is same, however, the side chain is absent or small in the mutant residue. Both mutant residues (Glycines and Cysteine) are hydrophobic as well as very flexible, and can disturb the required rigidity

of the protein at this position. This can result in loss of hydrogen bonds and/or disturb correct folding. Furthermore, the wild-type and mutant amino acids in both cases also differ in size and therefore can lead to loss of interactions with other proteins.

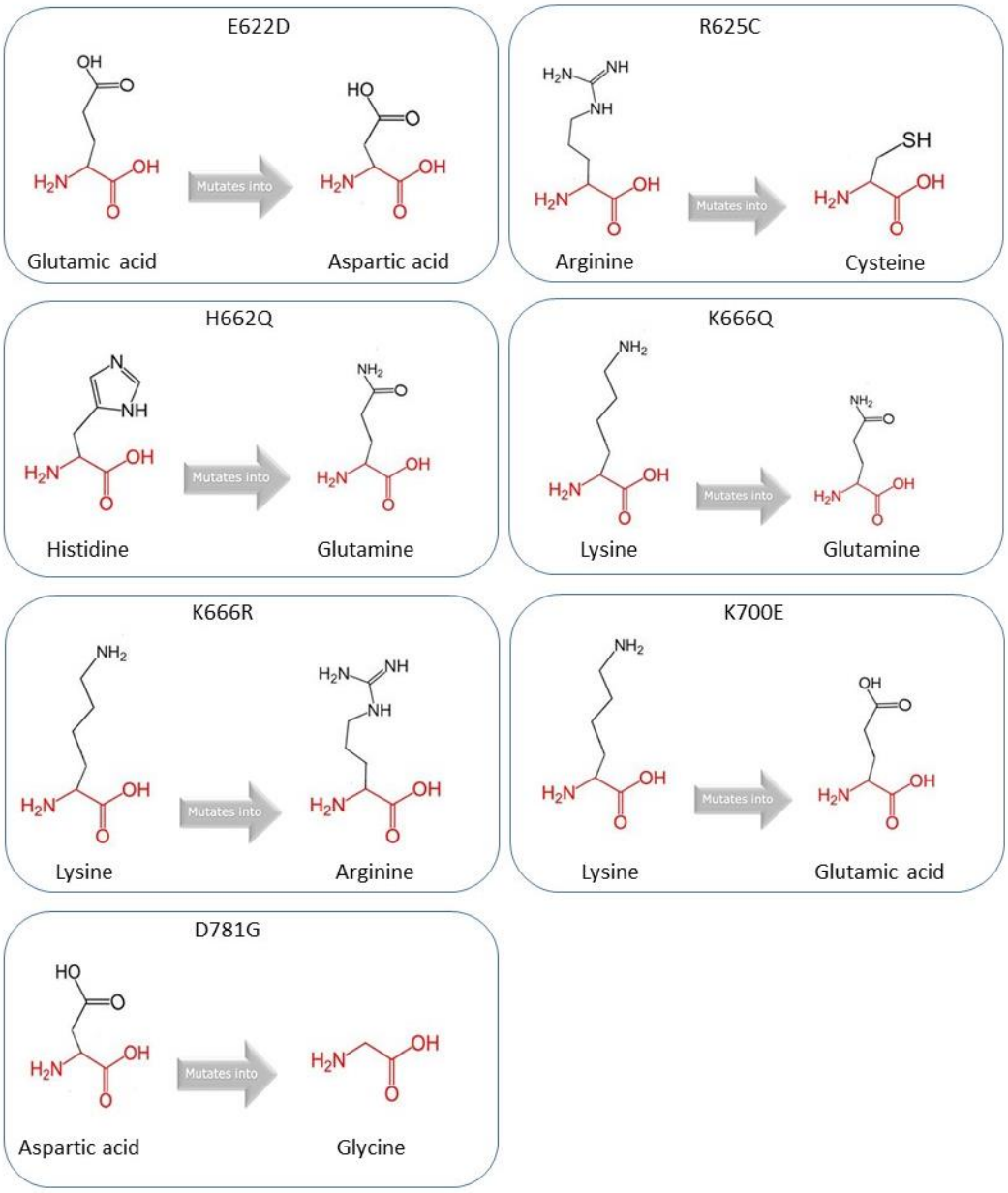


Figure 26: Chemical structure of the amino acids.

Each panel shows a structure of the respective wildtype and substituted amino acids observed in *SF3B1* mutant patients. Each amino acid has a backbone shown in red and a side chain shown in black.

3.1.4 *In-Silico* prediction for *SF3B1* mutations

Missense mutations that are disease-associated tend to alter residues that are evolutionary conserved. To explore the functional consequences of the observed amino acid substitutions in *in-silico*, we studied the potential degree to which the missense mutations can affect the function of the protein by using various algorithms such as Align-GVGD, SIFT and PolyPhen-2. These algorithms predict the impact of missense mutations on protein structure and function using algorithm-generated sequence alignments or manually curated alignments.

3.1.4.1 SIFT: Prediction of tolerated and deleterious *SF3B1* mutations

SIFT (Sorts Intolerant From Tolerant) algorithm predicts the consequences of an amino acid substitution on the function of the protein by considering the sequence homology and physical properties of amino acids. SIFT classifies mutations as “deleterious” or “neutral” using normalized probabilities calculated from the input multiple sequence alignment. Genomic variants at a position with normalized probabilities <0.05 are predicted deleterious while as a probability score of ≥0.05 is predicted as neutral or tolerated.

SIFT analysis revealed that bulk of the *SF3B1* mutations are likely to alter the function of the protein (Table 26). Of the 7 *SF3B1* mutations, majority of the mutations including the commonest amino acid change Lys700Glu were predicted to be deleterious with SIFT score of <0.05. Interestingly, Lys666Gln was shown to have a damaging effect (SIFT score 0), whereas Lys666Arg was predicted to have no effect on the protein function (SIFT score 0.17).

Table 26 : *In-Silico* prediction for *SF3B1* mutations.

		SIFT		PolyPhen-2		Align-GVGD	
Substitution	**Score**	**Prediction**	**Score**	**Prediction**	**GV**	**GD**	**Prediction**
Arg625Cys	0	Damaging	1	Probably damaging	0	179.53	Class C65
Asp781Gly	0	Damaging	0.999	Probably damaging	0	93.77	Class C65
Lys700Glu	0	Damaging	1	Probably damaging	0	56.87	Class C55
Lys666Gln	0	Damaging	1	Probably damaging	0	53.23	Class C45
Glu622Asp	0.05	Damaging	0.877	Possibly damaging	0	44.6	Class C35
His662Gln	0.01	Damaging	1	Probably damaging	0	24.08	Class C15
Lys666Arg	0.17	Tolerated	0.93	Possibly damaging	0	26	Class C25

3.1.4.2 PolyPhen-2: Prediction of *SF3B1* mutant protein structural and functional modifications

PolyPhen-2 algorithm is the latest tool created by the developers of the PolyPhen. Its novel elements include the set of predictive features, the alignment pipeline, and the probabilistic classifier based on machine-learning methods. PolyPhen-2 prediction is based on eight sequence-based and three structure-based predictive features that are selected automatically by an iterative algorithm. These include comparison of the features of the wild-type allele and the mutant allele. Essentially, PolyPhen-2 characterizes protein structural modifications caused as a result of amino acid substitutions which effect protein structure and function. PolyPhen-2 provides both qualitative prediction such as ‘probably damaging’, ‘possibly damaging’, ‘benign’ or ‘unknown’ and a score. Another useful feature is that the algorithm calculates a Bayes posterior probability that a given mutation is deleterious.

PolyPhen-2 analysis revealed that *SF3B1* mutations such as Lys700Glu, His662Gln, Lys666Gln and Arg625Cys were likely to have a damaging effect (probably damaging, PolyPhen-2 score 1) on the protein (Table 26). Likewise, Asp781Gly was predicted to be ‘probably damaging’ with a PolyPhen-2 score of 0.999. However, Glu622Asp and Lys666Arg mutations had a PolyPhen-2 score of 0.877 and 0.93, respectively.

3.1.4.3 Align-GVGD:

Another algorithm program namely Align-GVGD predicts consequences of variants based on a combination of Grantham Variation (GV), which measures the biochemical evolutionary variation at a particular position in the sequence alignment, and Grantham Deviation (GD), which measures the biochemical difference between the reference and variant amino acid. Align-GVGD uses a set of five criteria based on *GV* and *GD,* which groups variants ranging from the most pathogenic to least likely pathogenic. The program provides a series of ordered classifications ranging from the most likely deleterious 'C65' to the least likely deleterious 'C0' [deleterious 1 (GV = 0 and GD > 0), deleterious 2 (0 < GV ≤ 61.3 and GD > 0), neutral 1 (GD = 0), neutral 2 (GV > 61.3 and 0 < GD ≤ 61.3) and unclassified]. This algorithm has previously been used for developing in-silico models for a few clinically relevant tumour suppressor genes such as BRCA1 (Tavtigian et al., 2006) and TP53 (Mathe et al., 2006).

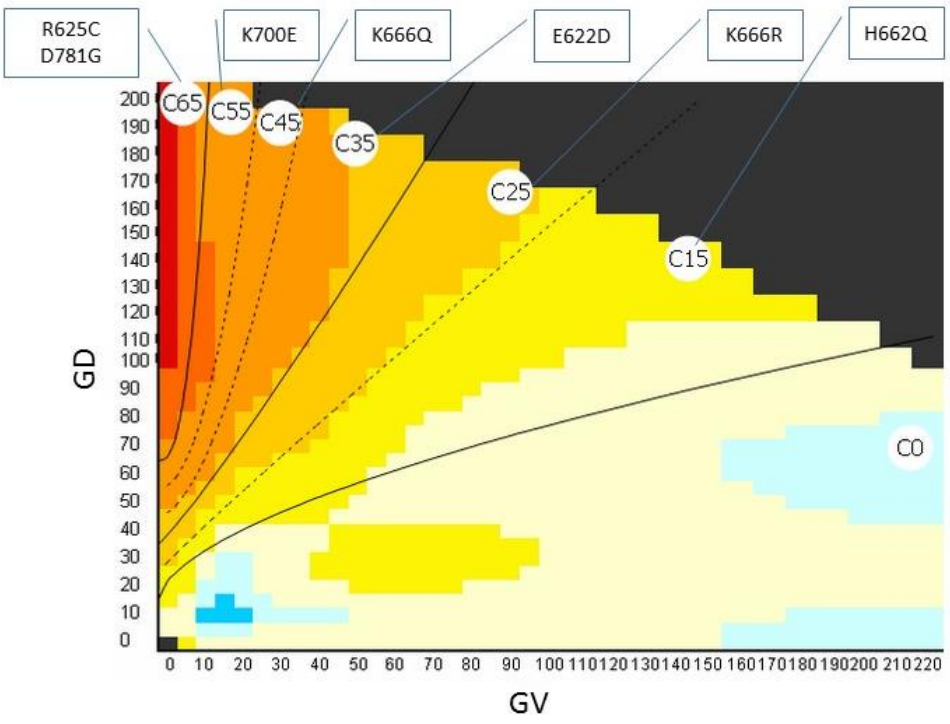


Figure 27: Align-GVGD algorithm prediction for SF3B1 mutations.

Align-GVGD uses a set of five criteria based on GV and GD, which groups variants ranging from the most pathogenic to least likely pathogenic (deleterious 'C65' to the least likely deleterious 'C0'). Figure shows all the *SF3B1* variants including the most frequent variant Lys700Glu fall between the deleterious regions C15 to C65.

Among the *SF3B1* mutations, Arg625Cys and Asp781Gly had the highest align-GVGD score (Class C65) and therefore likely to have deleterious effect on the protein (Table 26 and Figure 27). The most frequent *SF3B1* mutation Lys700Glu was also found to be highly deleterious (align-GVGD Class C55). Furthermore Glu622Asp, His662Gln and Lys666Arg had an align-GVGD score of C35, C15 and C25, respectively. These mutations were less likely to have any effect on the protein function.

Together, this data suggest that mutated *SF3B1* may affect the structure and functions of protein, in the form of mis-folding and intermolecular interactions, and play an important role in the disease susceptibility.

3.1.5 Exome sequencing reveals additional gene mutations in MDS-RS patients

Next, we sought to identify if there are additional somatic mutations affecting any specific pathway(s) in these *SF3B1* mutated patients. To investigate this, we selected our candidate mutations which passed out “mutation filtration” criteria (see methods section 2.2.4.6 and Figure 10 for more details). Between 4 to 16 well-supported candidate mutations were identified per individual exome and subjected to validation using targeted mutation approach and Sanger sequencing (see methods section 2.2.5, 2.2.6 and/or 2.2.7 for more details). Following the validation of the candidate mutations, our data revealed other somatic mutations (Table 25) in MDS-RS patients which included the known myeloid-related gene targets in myelodysplasia with similar mutation frequencies as previously reported, therefore indicating an acceptable sensitivity of the current study. The mutations of the known myeloid gene targets, however, accounted for only 16% of all detected mutations (n=56), and the remaining confirmed novel 36 mutations in the previously unreported genes.

Among the known myeloid-related genes, epigenetic modifier gene mutations were present in 8/12 MDS-RS patients (*DNMT3A*, n=6; *TET2*, n=2). Interestingly, single epigenetic modifier gene mutations were observed coinciding with *SF3B1* aberrations. Moreover, mutual exclusivity could exist among genes of the same functional categories, because no concurrent presence of gene mutations of splicing complex or epigenetic modifiers were detected in the same patient. Like *SF3B1* mutations, all the epigenetic modifier mutation were heterozygous (MAB 41% to 50%). However, one

patient had a *DNMT3A* R882H mutation with an MAB of 75%, indicating a homozygous state or loss of heterozygosity (LOH) at chromosome 2p23.3. *JAK2* which is a cell signalling protein was mutated (V617F, MAB 43%) in one patient (MDS1). Of particular note, none of the 'oncogenic' gene mutations such as cell signalling proteins (*KRAS*/*NRAS*, *C-Kit*, and *PTPN11*) or tumour suppressor proteins (*TP53*) appear to be mutated in MDS-RS patients, therefore indicating a possible mutual exclusivity.

WES revealed additional 36 patient specific novel mutations in $CD34^+$ cells (Table 25). These mutations include ones in previously uncharacterized zinc finger proteins, and mutations in factors which suggest defects in metabolic pathways (*SLC7A2, ANXA7, IKBKE, AGXT2L1, MC2R, ATP1A4, TESC, SPATA13, ABCB11, NAGA, NMNAT2, TLL2, RETNLB*), cell proliferation (*PPM1D*), DNA maintenance (*CEP110, AFF3, G3BP1, HIST1H2AC*), transcription regulation (*MCM3AP, CAMTA1, NR0B2*) and cytoskeletal organization (*FITM2, ATG9B, PCNT, PKHD1, GPR112, FGD1, MYH14, PTPDC1, TH, FBLIM1* ; Table25). However, mutations in the haem biosynthesis pathway enzymes (for example, *ALAS2*, *ALAD*, *FECH* and *UROD*) and globin genes (for example, *HBQ1*, *HBA2*, *HBB* and *HBA1*) were not detected in our patient cohort. Furthermore, *in-silico* analysis of all the gene mutations demonstrated that some of the gene aberrations disrupt the function of the respective protein while others are benign (Table 25). However, more analysis is needed to unravel the role of these mutations in *SF3B1* mutant patients.

3.1.6 Architecture of Clonality in MDS-RS patients

Previous studies have established MDS as clonal diseases by using X-linked polymorphic markers and molecular cytogenetics (Boultwood and Wainscoat, 2001; Janssen et al., 1989; Mongkonsritragoon et al., 1998). However, recently the use SGS and various bioinformatics software's have led the researchers to infer the clonal architecture in the total bone marrow samples through the identification of clusters of co-existing mutations with similar MABs. Therefore, based on the MAB of the mutated genes obtained from the genome sequences (WES data) of confirmed SNVs, we sought to identify the clonal architecture of the bone marrow compartment in these patients. We hypothesized that MAB should be 50% for heterozygous mutations or 100% for homozygous mutations, if mutations exist in all tumour cells, but it is often

lower because of the admixture of non-tumour elements in clinical samples. With the sensitivity of this approach, the peak of greater MAB value should represent the founding clone, whereas the peak(s) with smaller MAB may represent subclone(s), although very small subclone(s) could be under the limit of detection with this sequencing approach. Another major problem with using this approach is that we won't be able to determine confidently whether a mutation is present in homozygous or heterozygous state.

Based on this approach, in our patient cohort *SF3B1* mutations are present in-between 45% to 50% MAB, indicating that these mutations are present in all the bone marrow cells in a heterozygous state. Similarly, majority of the other gene mutations including the cell signalling proteins (such as *JAK2*, MDS1 36% MAB) and patient specific mutations (such as *TLL2*, MDS2 7% MAB; *ANAX7*, MDS10 26% MAB) are present at <50% MAB, and likely to be heterozygous mutations (Figure 28). In contrast, some of the gene mutations such as *DNMT3A* (MDS11, 75% MAB), *MC2R* (MDS8, 63% MAB) are present at a MAB of >50%, indicating that these aberrations are homozygous mutations (Figure 28 and Table 25) or the presence of LOH at the genomic locus, and these mutations are present in a fraction of the bone marrow cells. Therefore, it is difficult to conclude the clonal composition of the bone marrow compartment in some patients where gene mutations are present at different frequencies. However, in majority of our patients *SF3B1* mutation appears to be the initiating mutant clone due to its higher abundance in the bone marrow cells while as other gene mutation are likely to be the subclone(s). Given that our data suggests that *SF3B1* mutations are present in all the bone marrow $CD34^+$ cells, therefore they should reflect the aberrant cellular activities at very high levels of disease involvement along with the hematopoietic cell proliferation/ differentiation.

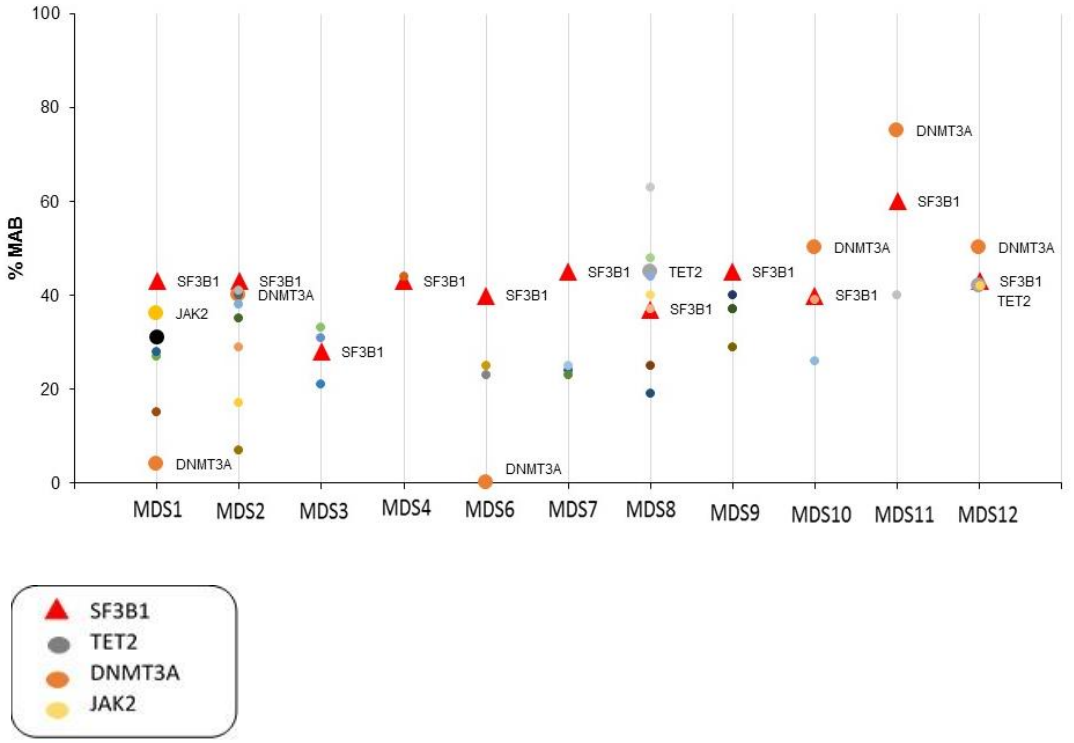


Figure 28: Prediction of order of genetic lesions in MDS-RS patients.

Patients with one or more candidate driver mutation were identified by whole-exome sequencing (followed by confirmation by target mutational sequencing). Vertical lines represent MDS-RS patients. SF3B1 and known myeloid related gene mutations are shown by labelled coloured circles. Other coloured circles represent patient specific mutations detected through whole-exome sequencing (see table 25 for more details).

3.1.7 *SF3B1* mutation persists throughout the disease progression

Leukemic transformation in patients with RARS is a very rare event. One of the patient (MDS54) from our cohort with RARS and *SF3B1* mutation evolved to AML at 4 months after diagnosis. To investigate whether *SF3B1* mutation persists throughout the disease evolution, relevant sequential samples were screened using SGS platform. 454-Roche sequencing analysis of the samples taken from patient at the time of diagnosis revealed *SF3B1* mutation with a mutant allele burden of 39% (Figure 29). At the time of transformation to AML, *SF3B1* mutation was maintained at the similar level (Figure 29). Following intensive chemotherapy (Fludarabine, Cytarabine and Filgrastim, FLAG), patient attained a transient morphological remission but relapsed rapidly with the *SF3B1* MAB remaining constant (MAB 41%) throughout the disease duration, which as a heterozygous mutation would occur in a majority of sample cells. Therefore, our data suggests that *SF3B1* mutation in this patient originates from the

early haematopoietic stem cells which is not eradicated by the intensive chemotherapy and persists throughout the disease progression.

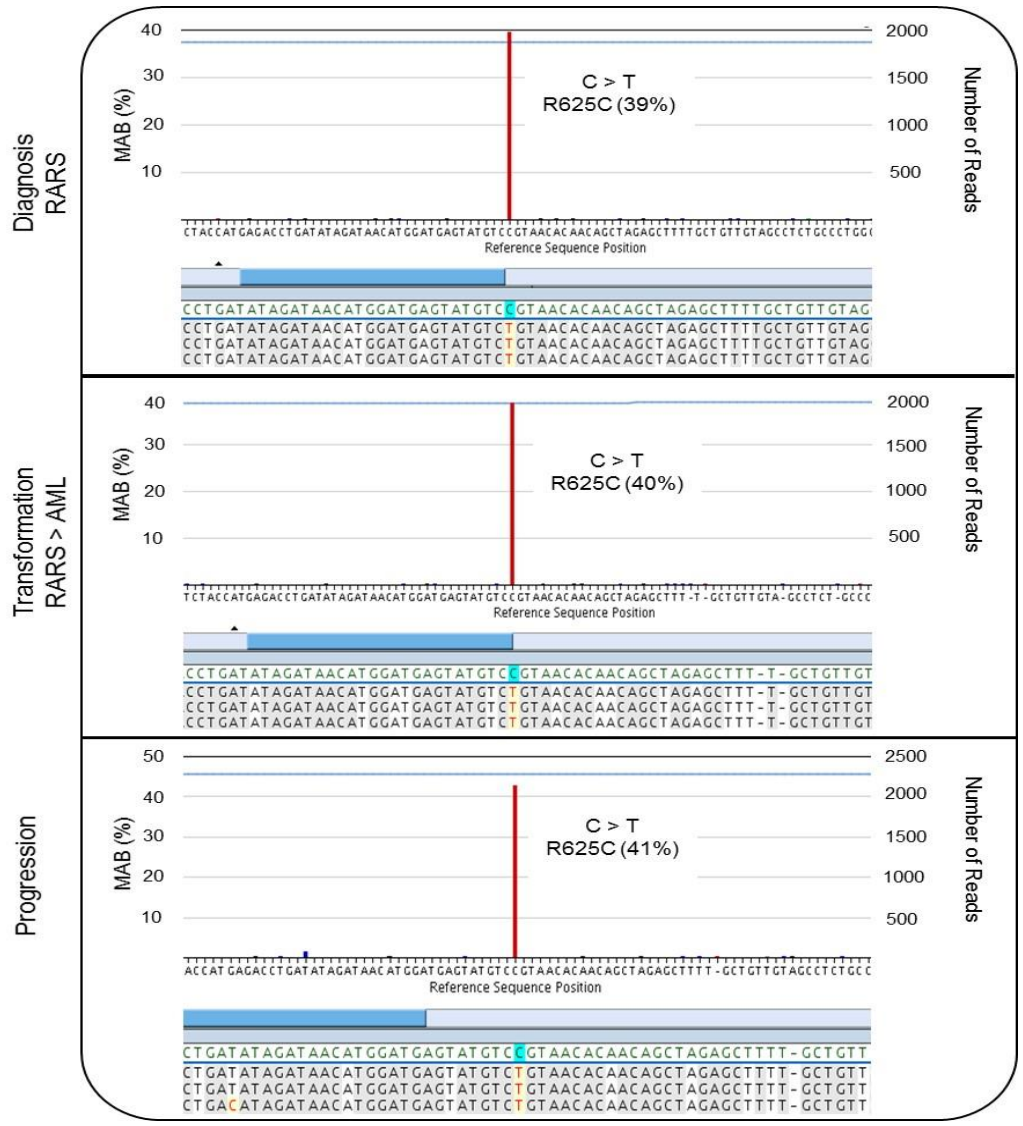


Figure 29: *SF3B1* mutation persists during the disease progression.

Samples were obtained at diagnosis, transformation and progression. Distribution of generated amplicon sequence reads per patient. For an exemplary patient, the read length (x-axis) and number of generated sequence reads (y-axis) is represented. Mutation burden is on the (y axis). *SF3B1* mutation burden was maintained at the same level from diagnosis, at progression to AML and at time of remission following intensive chemotherapy.

3.1.8 *SF3B1* mutations have no impact on the clinical outcome of MDS patients with ring sideroblasts

Patients with an *SF3B1* mutation had a higher platelet count as compared to patients who did not have an *SF3B1* mutation (mean, mutant 335x10^9/L vs wildtype 85x10^9/L; P=0.003) as well as high neutrophil count (mean, mutant 3.1x10^9/L vs wildtype 1.5x10^9/L; P=0.059; Figure 30). There was no significant difference in haemoglobin levels (mean, mutant 9.148g/dl vs wildtype 9.533g/dl; P=0.059; Figure 30) and ferritin levels (mean, mutant 1308µg/L vs wildtype 1312µg/L; P=0.61) between *SF3B1* mutant vs wildtype patients. Univariate analysis revealed that there was no difference in overall survival (OS) or leukaemic transformation when comparing patients with or without *SF3B1* mutation (Figure 31).

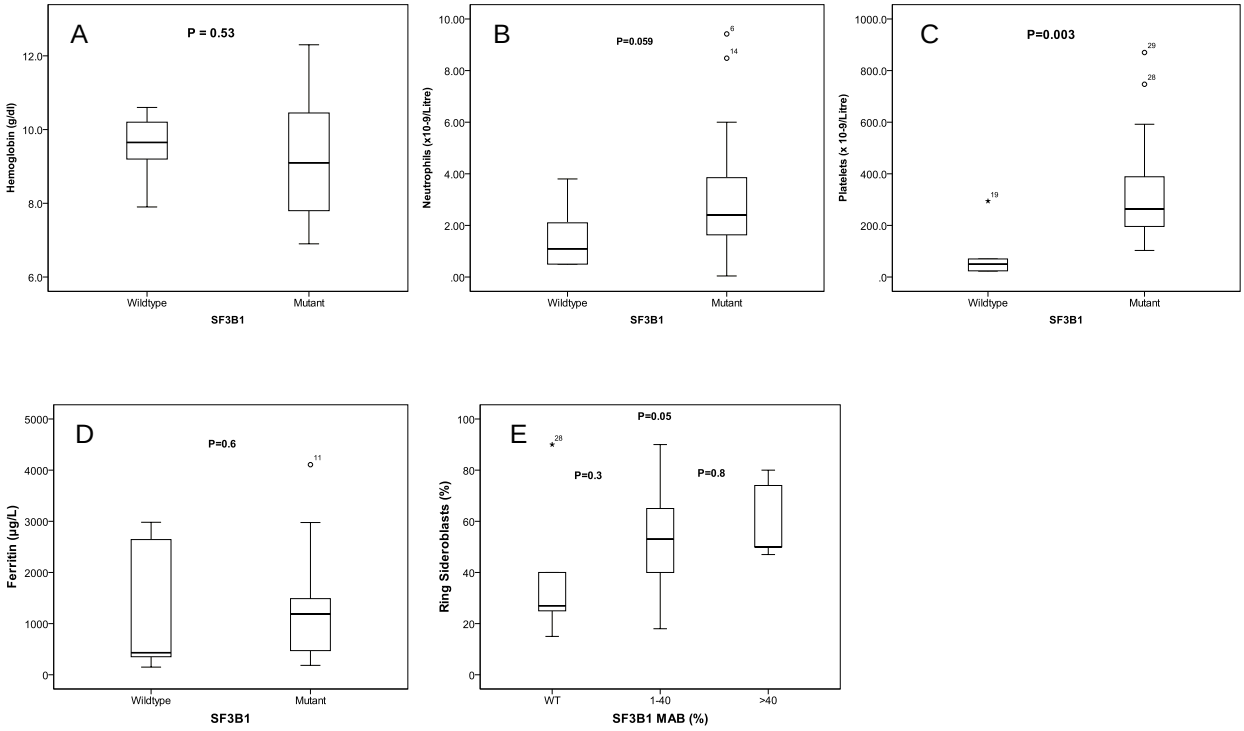


Figure 30: Relationship between *SF3B1* mutation status and blood/bone marrow findings.

Panel A-D shows the box-and-whisker plots for haemoglobin levels, neutrophil counts, platelet counts and ferritin levels. The circles above and below the box-and-wisker plots indicate outlier data for individual patients. Panel E shows the box-and-wisker plot for *SF3B1* MAB and percentage of ring sideroblasts. Values for percentage of MAB are grouped here in 3 arbitrary categories: *SF3B1* WT (n=5), *SF3B1* MAB 1-40% (n=9), and *SF3B1* MAB >40% (n = 11). MAB- Mutant allele burden.

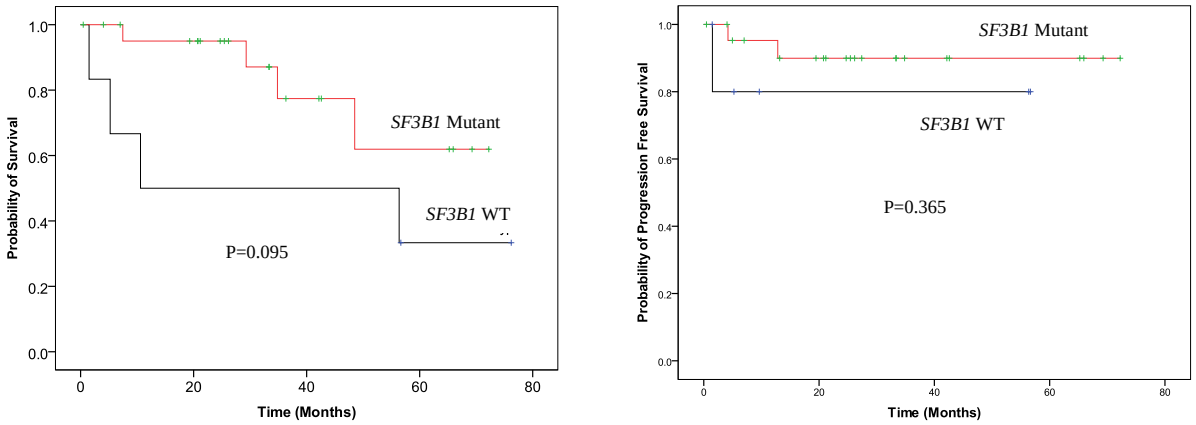


Figure 31: *SF3B1* mutations have no impact on the overall survival of the patients.
No significant difference in overall survival or leukemic transformation when comparing patients with or without *SF3B1* mutation. The P values are derived from Kaplan-Meier analysis with a log-rank test and median survival times with 95% confidence intervals.

Data for the estimation of the ring sideroblasts using consensus criteria (Mufti et al., 2008) was available for 25 of the 29 MDS patients. Ring sideroblasts were identified at variable percentages with a mean of 53.16% (range, 15%-90%). Among the patients with *SF3B1* mutations, there was no significant difference in the percentage of ring sideroblasts and the *SF3B1* MAB (Figure 30). However, patients with MAB of >40% had higher percentage of ring sideroblasts as compared to patients with wildtype *SF3B1* (Figure 30). Of 23 patients with *SF3B1* mutations, 2 patients had 15-29% ring sideroblasts, 9 patients had 30-50% ring sideroblasts and 9 patients had 61-90% ring sideroblasts (percentage of ring sideroblasts was not available for 3 patients). There was no significant difference between the percentage of ring sideroblasts vs the MAB within the *SF3B1* mutant group (Figure 31).

3.2 Transcriptome analysis of progenitor cells from *SF3B1* mutant patients

To date, question remains as how *SF3B1* mutations can give rise to the phenotypic changes observed only in the erythroid progenitor cells in MDS-RS patients. Therefore, we set out to identify the downstream effects of *SF3B1* mutations in acquired sideroblastic anaemia patients which will help us to identify the aberrant

pathway(s) or gene(s) responsible for '*SF3B1* specific' sideroblastic phenotype. Bone marrow $CD34^+$ progenitor cells and $CD71^+CD235^+$ erythroid progenitor cells were isolated using BD FACS Aria SORP (San Jose, CA; See methods section 2.2.8 for full details). Three MDS-RS patients in whom *SF3B1* mutations were detected in total bone marrow cells were used for this analysis (Table 27). In addition, one acquired sideroblastic anaemia patient with wildtype spliceosome genes was used as a control (Table 27). The cells of interest were isolated by the cell surface markers (CD34, CD71 and CD235) as described in the methods section 2.2.8. After sorting, the purity of all the cell fractions was checked and a sorting purity of ≥97% was routinely obtained.

Table 27: Characteristics of the patients studied for Gene expression/Exon-array analysis.

Sample ID	Status	Gender	WHO subtype	Karyotype	Ring Sideroblasts (%)	*SF3B1* Mutation (Red=mutant allele)	Other known Common Myeloid gene mutations
MDS1	MDS-RS	M	RCMD-RS	46,XY	50%	K700E (43%)	JAK2 (V617F, 36%)
MDS4	MDS-RS	F	RARS	46,XX	24%	K700E (40%)	None
MDS6	MDS-RS	M	RCMD-RS	46,XY	65%	E622D (40%)	DNMT3A (R736H, 8%)
MDS28	MDS-RS	F	RARS	46,XX	Unknown	WT	ASXL1 (G643V, 32%)

Following the cell separation, gene expression/Splicing array analysis was perform using Affymetrix GeneChip Human Transcriptome Arrays (version 2.0; HTA2). These high-resolution arrays contain >6.0 million distinct probes for coding and non-coding transcripts. Furthermore, 70% of the probes cover exons for coding transcripts while as the remaining 30% of probes target exon-exon splice junctions and non-coding transcripts. This Affymetrix array is the only commercially available microarray for expression profiling at both gene as well as exon levels; although it provides accurate assessments of gene expression.

The raw images from the DAT file generated from Affymetrix array probes were converted to CEL files. Initial data analysis (including normalization) was performed using Affymetrix expression console (Figure 21). The quality of the data was checked manually by looking the QC matrix generated by the Affymetrix expression console (Figure 32). Principle component analysis (PCA) divided the samples into two groups ($CD34^+$ and $CD71^+CD234^+$ groups) as shown in Figure 33.

Affymetrix TAC v2.0 was used to perform one-way ANOVA analysis to detect differentially expressed gene as well as differential splicing index between the *SF3B1* mutant vs *SF3B1* wildtype patient samples ($CD34^+$ cells and $CD71^+CD235^+$ cells, Table 28, Appendix Table 15 and Appendix Table 16). The data was filtered further using a FDR <0.1 (unadjusted P value) and a fold-change cut-off of ≥2 (or Splicing Index ±2) to create a list of significantly up or down regulated genes and differentially splicing gene transcripts.

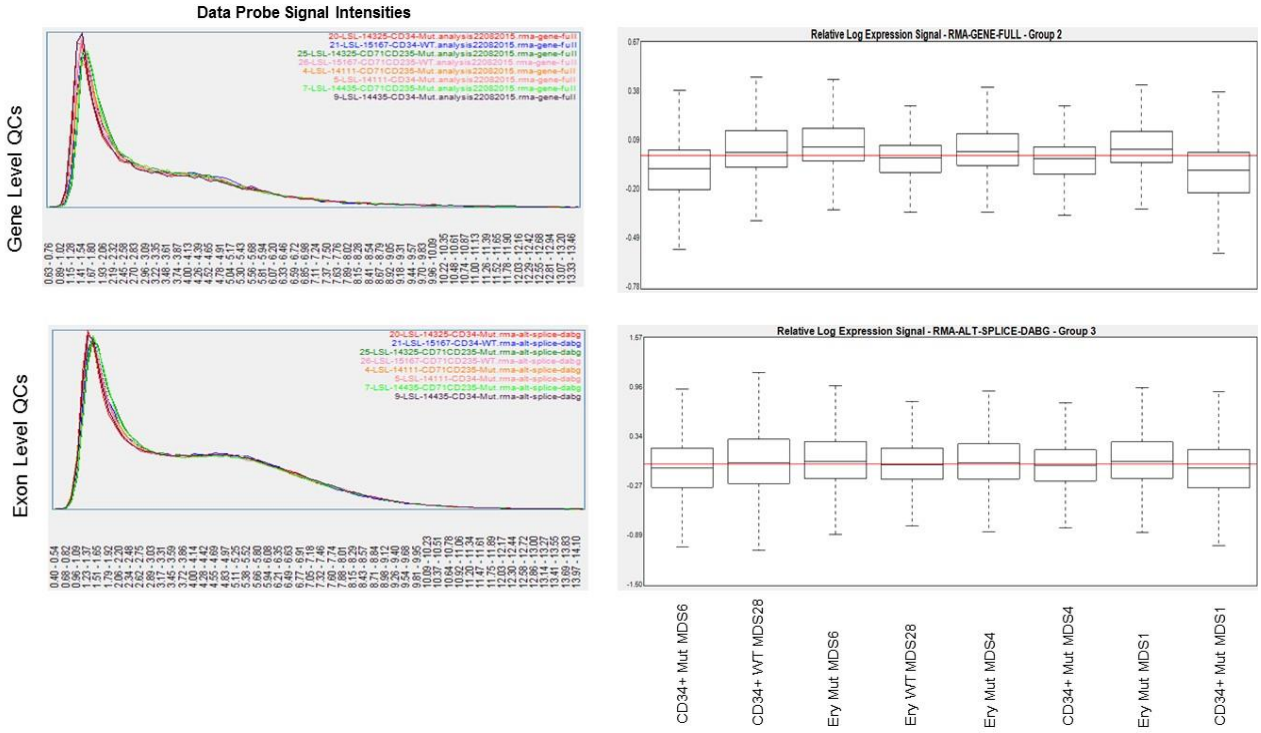


Figure 32: QC plots produced for normalized arrays for each sample.

(A)Shows the data probes intensities for each array. (B) Shows the relative expression signals for each array after the RMA-gene and RMA alternative splicing analysis. Both figures demonstrates that the intensity signals generated by all (with *SF3B1* Wildtype and *SF3B1* mutant samples) the arrays is similar.

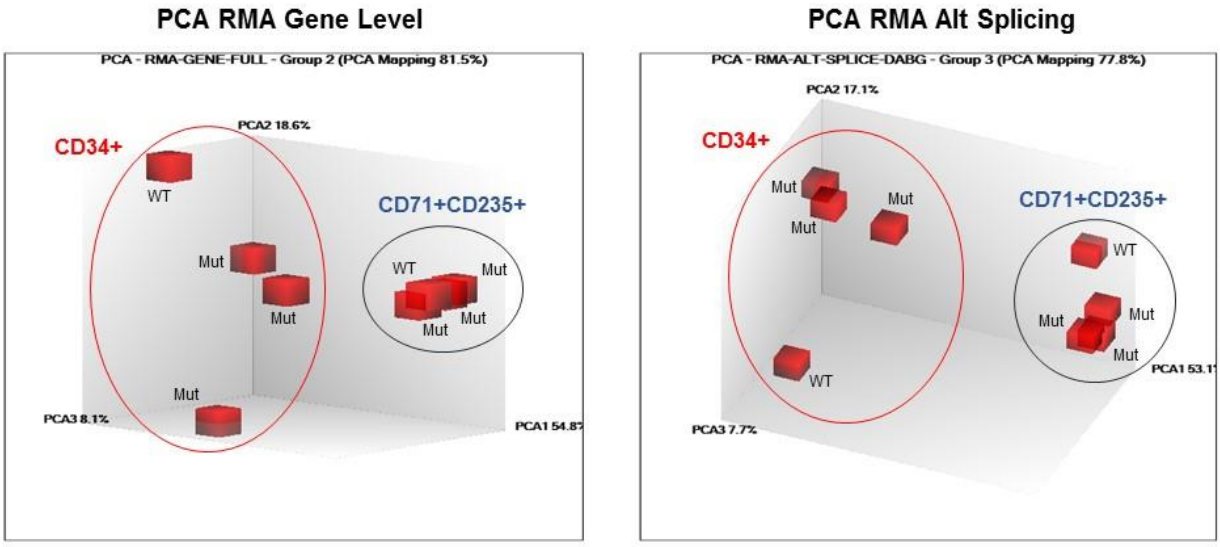


Figure 33: Principle component analysis (PCA) preformed using Affymetrix Console Software.

The PCA graphs of the global gene expression data (A) and alternative splicing data (B) were computed using Affymetrix console software. Bone marrow cells were grouped into two groups i.e. $CD34^{+}$ and $CD71^{+}CD235^{+}$.

A list of genes was generated (Table 28) after the preliminary analysis i.e. normalization and annotation. However, no significant difference in the gene expression profile was observed between the *SF3B1* mutant vs *SF3B1* wildtype groups (Appendix Table 15 and Appendix Table 16) after applying the filtration criteria. In addition, we did not observe any significant difference in the splicing index's between the two groups (Appendix Table 17 and Appendix Table 18).

Table 28: Details of the data generated following HTA 2.0 array.

Data analysis by Affymetrix transcriptome analysis console. ** Applying data filtration as described in methods chapter.

Gene Level analysis		
	$CD34^+$	$CD71^+CD235^+$
Total no of Genes	67528	67528
Coding Genes	44699	44699
Non Coding Genes	22829	22829
Differentially expressed		
Coding genes	389	296
Non-Coding	139	114
Up-regulated Genes		
Coding	263	235
Non-Coding	103	104
Down-regulated Genes		
Coding	126	61
Non-Coding	36	10
Genes after data filtration*	0	0
Alternative Splicing Analysis		
	$CD34^+$	**$CD235^+$**
Total no of Genes	67539	67539
Coding Genes	44710	44710
Non Coding Genes	22829	22829
Coding		
Genes expressed in both conditions	29302	29916
Genes expressed genes to indicate alternative splicing	7143	3871
Genes expressed in *SF3B1* WT but not in Mutant	1672	4826
Genes expressed in *SF3B1* Mutant but not in WT	4680	1340
genes not expressed in WT/Mutant conditions	9056	8628
Non-Coding		
genes expressed in both conditions	12191	13226
Genes expressed genes to indicate alternative splicing	506	240
Genes expressed in *SF3B1* WT but not in Mutant	1340	2832
Genes expressed in *SF3B1* Mutant but not in WT	2917	1078
genes not expressed in WT/Mutant conditions	6381	5693
Genes after data filtration**	0	0

3.3 Discussion

RARS is an acquired form of MDS mainly characterized by an initial phase of erythroid hyperplasia with an excess iron accumulation in the mitochondria of erythroblasts which subsequently leads to ineffective erythropoiesis (Cazzola and Invernizzi, 2011). In a significant proportion of patients, this is followed by a phase of marrow failure, with or without the emergence of leukemic blasts. Over the past few years, various genetic factors have being associated with RARS (Boultwood et al., 2008; Cazzola and Invernizzi, 2011; Malcovati et al., 2009; Nikpour et al., 2010). However, the mechanisms underlying the pathogenesis of RARS and the cause of the unusual pattern of iron deposition within the erythroid cells remains yet to be determined.

In order to understand the mechanism underlying the pathogenesis of RARS, WES was performed on $CD34^+$ cells from 11 MDS patients with ring sideroblasts (MDS-RS) as such progenitors are likely to provide the reservoir of myeloid mutations, as has been previously reported (Anand et al., 2011; Smith et al., 2010). Our WES data revealed recurrent mutations of the *SF3B1* gene in 11/11 cases, which as such, demonstrated the unquestionable power of second-generation sequencing (SGS) technologies in cancer research. No *SF3B1* mutation was detected in one patients with congenital sideroblastic anaemia. However, a constitutional mutation in *ALAS2* gene which has been previously (Camaschella, 2009; Cotter et al., 1999; Kaneko et al., 2013) linked with sideroblast phenotype was detected in this patient. Follow-up targeted sequencing in additional 18 MDS patients with ringed sideroblasts identified acquired *SF3B1* mutations in 86% (6/7) RARS, 100% (1/1) RARS-T, 44% (4/9) RCMD-RS and 100% (1/1) AML with a preceding RARS. All the mutations were confirmed as acquired somatic mutations by their absence in DNA isolated from paired skin biopsy or $CD3^+$ T-cells. Two studies published at the same time reported similar data showing the presence of *SF3B1* mutations in MDS patients with ring sideroblasts (Papaemmanuil et al., 2011; Yoshida et al., 2011b). In fact, *Yoshida et al* unmasked a complexity of novel mutations affecting multiple but distinctive components of the splicing machinery in approximately 45% to 85% of MDS patients depending on the disease subtypes.

Precise removal of pre-mRNA introns, a critical aspect of gene expression is catalysed by a highly dynamic protein complex, the spliceosome (Kuwasako et al., 2008). *SF3B1* located on chromosome 2q33.1, is an essential component of the U2 snRNP which along with other snRNPs forms the main building block of the major spliceosome, responsible for removing the vast majority of pre-mRNA introns (Kuwasako et al., 2008). Sf3b1 lies within the U2-snRNP spliceosome catalytic site, and has an important role in the pre-spliceosome assembly and splicing catalysis. Sf3b1 is composed of two regions: the *N-t*erminal *d*omain (Sf3b155 NTD) and the C-terminus. The N-terminal of Sf3b1 is unstructured; less conserved and contains the binding sites for splicing associated factors U2AF65, p14, Cyclin E, and Nipp 1 (Wahl et al., 2009). However, the C-terminus of *Sf3b1* is highly conserved among different species, including human, mouse, schizosaccharomyces pombe, caenorhabditis elegans and Xenopus (Isono et al., 2001). All the *SF3B1* mutations detected in our patient cohort were present in this highly-conserved protein c-terminal. This part of the protein contains 22 tandem helical repeats (HEAT motifs) that form rod–like, helical structure and wrap around the U2 snRNP, therefore providing the stability to the whole U2 snRNP complex of the major spliceosome. It is worth to note that majority of the mutations were present in the sixth HEAT motif while as some were also present in the fourth and fifth HEAT motif. Each of the HEAT motif contains repeats of ≈47 amino acid residues. The structural studies of Sf3b have revealed that these HEAT motifs wrap round the U2 snRNP and are thought to stabilize the structure of the U2 snRNP. The conformation of Sf3b bound to the U11/ U12 di-snRNP differs from that of isolated Sf3b and upon integration into the di-snRNP, Sf3b rearranges itself into a more open form (Golas et al., 2005; Stark and Luhrmann, 2006). The candidate domain within the whole SF3b complex that drives this rearrangement is likely to be Sf3b1 HEAT motifs itself. This observation is based on the similar reports of the changes in the curvature of the proteins with HEAT repeats which have been previously reported in protein complexes such as importin-β. Importin-β HEAT motifs upon binding to one of its interaction partners importin-α or Ran changes its conformation (Cingolani et al., 1999). The sixth HEAT motif in which majority of the mutations are present, lies at the hinge of the shell-like structure which plays a major role in this rearrangement process of Sf3b1 (Golas et al., 2005). There is an increasing evidence suggesting that initial splicing occurs as the nascent RNA molecule is being transcribed, an integrated

process in which a continuous cross-talk exists between the spliceosome and the proteins involved in the initiation, elongation, and termination stages of the transcription machinery (Chen and Manley, 2009). Therefore, knowing the critical function of the Sf3b, the likely consequence of these mutations would be the impairment of the switching mechanism by which Sf3b1 changes its conformation and interacts with other spliceosome components, therefore leading to the disruption of the splicing itself. On the other hand, it might have a direct effect on the interaction of Sf3b1 or spliceosome with the active transcriptional machinery.

Functional characterization of *SF3B1* mutations and their contribution to myelodysplasia is still unclear. Sf3b1-/- homozygous null mutant mice have been shown to die very early, around the 16- to 32- cell stage (Isono et al., 2005), therefore reinforcing the importance of *SF3B1.* It is worth to note that all the *SF3B1* mutations are heterozygous substitutions and none of aberrations shown in this study or reported elsewhere are homozygous substitutions or frameshift and splice-site aberrations. Therefore these aberrations are likely to produce a structurally intact but functionally impaired Sf3b1 protein with perhaps an altered gain-of-function protein. This is also reflected from our i*n-silico* mutational analysis which suggests that *SF3B1* mutations will likely have an effect on the function of the protein as has also been suggested by others (Papaemmanuil et al., 2011).

A connection between pre-mRNA processing and cell cycle progression has been implicated by various studies including a study by our group (Orr et al., 2012). In a transient experiment, Steve and colleagues showed reducing the expression of SF3B2 or SF3B4 prevented T cells entering the cell cycle (Orr et al., 2012). Furthermore, i*n-vitro* knockdown experiments have linked loss of Sf3b1 function with the aberrant cell cycle characteristics, cell cycle arrest and increased apoptosis, where aberrant splicing of cell-cycle genes has been noted (Kaida et al., 2007; Yoshida et al., 2011a). In support of this, a study by *Jonathan et al* reported that 5 novel splicing isoforms of *CDC25C* were found in MDS patients, and the splicing patterns were generally distinct in patient samples compared with healthy controls. One of the novel isoforms *CDC25C-6* was found to be present in 58% of the patients. The highly conserved *CDC25* phosphatases is a critical component of cyclin/CDK regulation. *CDC25A*, is involved in initiation of the S phase of cell division, while the other two, *CDC25B* and

CDC25C, regulate the G2/M transition (Caudill et al., 2008). This study provides a possible link between the aberrant splicing of cell cycle gene and the MDS pathophysiology.

Somatic *SF3B1* mutations are found almost (up to 85%) in all MDS patients with ring sideroblasts (RARS, RCMD-RS and RARS-T), therefore suggesting a causal relationship between mutation and its downstream effects resulting in the formation of ring sideroblast. Although, the majority of the *SF3B1* aberrations are dominant heterozygous mutations, however, the percentage of ring sideroblasts in these patients ranges considerably from 20% to 95%. Interestingly, various studies (Cazzola et al., 2012; Damm et al., 2012c; Papaemmanuil et al., 2011) including ours have found a relationship between the *SF3B1* mutations and the percentage of ring sideroblasts, however, this does not explain the huge phenotypic (i.e. ring sideroblasts percentage) variability in these patients. A recent study by *Visconte et al* has implicated the role of *SF3B1* aberrations in the formation of ring sideroblasts in MDS (Visconte et al., 2012b). The author found that meayamycin which is a pharmacologic inhibitor of the Sf3b complex induces the formation of ring sideroblasts in the healthy bone marrow cells (Visconte et al., 2012b). In agreement with this, transient siRNA-mediated silencing of *SF3B1* elsewhere (Nikpour et al., 2012) resulted in reduced expression of *ABCB7* in K562 cells. Down regulation of *ABCB7* led to markedly reduced erythoid growth in *in-vitro* with accumulation of mitochondrial ferritin in immature RBCs. Furthermore, normalization of *ABCB7* expression in these cells rescued the RARS phenotype. In the same study, 11/13 studied RARS patients with *SF3B1* mutation had aberrant *ABCB7* exon usage as compared to normal bone marrow cells (Nikpour et al., 2012), therefore suggesting a link between the two genes and the RARS disease phenotype. On the other hand, recently mitochondrial transporters SLC25 (SLC25A37 and SLC25A38) and ALAD genes were implicated in the pathogenesis of ring sideroblast phenotype in RARS (Del Rey et al., 2015). In contrary to this, our gene expression data did not reveal any obvious downstream target genes (such as mitochondrial pathway genes or haem biosynthesis genes) specific to *SF3B1* mutations. This could be due to the fact that the net downstream effect of *SF3B1* mutations is very subtle. Due to the high turn-over of the erythroid cells in the body and hence higher requirement of the Sf3b1 protein, this subtle effect

on the downstream “gene(s) or pathway(s)” is enough to cause the phenotypic changes in erythroid cells but are not detectable with the current available technology. However, the difference in the results between our study and other published studies could be due to the fact that we were trying to identify the ‘SF3B1 mutation’ specific effect (by comparing *SF3B1* mutated vs *SF3B1* wildtype patients, both groups with ring sideroblasts) while as other published studies were focussing on the ring sideroblastic phenotype (by comparing *SF3B1* mutated patients with ring sideroblasts vs *SF3B1* wildtype non-sideroblastic anaemia MDS subgroups). In addition, different coexisting mutations present alongside *SF3B1* mutations might also be influencing the outcome of gene expression data coming from different studies. However, whether ‘SF3B1-ABCB7’ relationship forms basis of the pathophysiology of RARS remains to be determined and will require detailed biochemical studies of the mutant protein. Fundamentally however, the influence of such *SF3B1* mutations is not just in myeloid tissue and RARS, but has also been observed in chronic lymphocytic leukaemia (CLL) and lymphoid tissue (Greco et al., 2012; Wang et al., 2011), suggesting that genetic background plays an important role in the functional manifestation of these aberrations.

RARS patients generally have a relatively benign clinical course with a small fraction of patients progressing to AML. All the patients with *SF3B1* mutations had normal cytogenetics except one patient with trisomy 8. There was no difference in overall survival or leukemic transformation when comparing patients with or without *SF3B1* mutation. We also have found that MDS-RS patients with *SF3B1* mutations have higher platelet counts and fewer bone marrow blasts.

MDS has long been considered an oligoclonal disorder with the abnormal clones originating from the early hematopoietic stem cells. Analysis of the clonal spectrum by means of the MAB analysis of the coexisting mutations can shed new light into the order (or sequence) of these mutations in the bone marrow stem cell compartment. The fact that *SF3B1* mutations were detected in the $CD34^{+}$ progenitor bone marrow cells in all the screened patients suggests that these mutations occur in the early haematopoietic stem cells. This is also evident from the bioinformatics clonal analysis performed on these patients where in majority of the cases, *SF3B1* mutations are detected in majority of the cells as suggested by the highest MABs. Additional mutation

(such as *ASXL1*, *DNMT3A*) detected in these patients are acquired downstream of the *SF3B1* mutant clones. Furthermore, mutation analyses performed on serial samples in one patient who was initially diagnosed with *RARS* and subsequently transformed to AML, also provides us with evidence that *SF3B1* mutations are an early ancestral events initiating from the stem cells. In this case, the *SF3B1* mutant clone survives various treatments during the course of the disease and is detected at the time of remission as well as progression. The contribution of *SF3B1* mutations in disease transformation or progression to AML may be limited, but they may instead provide a favourable environment or sufficient pressure for other more destabilizing mutations to occur.

Our WES data revealed additional novel patient specific mutations along with *SF3B1* mutations. These include mutations in genes involved in epigenetic machinery, cell proliferation, DNA maintenance and cytoskeletal organization. However, we did not find any mutation in genes involved in haem synthesis pathway and mitochondrial ion transport which is consistent with the previous findings (Papaemmanuil et al., 2011; Yoshida et al., 2011b). Among the known myeloid gene mutations, *DNMT3A* was frequently present alongside with *SF3B1* mutations, perhaps suggesting a possible link between the two genomic abnormalities. Interestingly, only single epigenetic modifier mutations (either *DNMT3A*, *ASXL1* or *TET2*) seemed to co-occur with *SF3B1* mutations, therefore highlighting the importance of genetic background and how particular mutation combinations may attain an acceptable biochemical equilibrium and stable disease. The phenotypic and/or genotypic variations between patients with *SF3B1* mutations may either be a function of the different *SF3B1* mutation, or reflect additional mutation(s) present in the MDS clone in these cases.

In summary, the newly identified mutation in *SF3B1*, found in a significantly high proportion of MDS patients with ring sideroblasts, provides a possible causative link to the typical phenotypic ring sideroblasts observed in this disease subtype. However, questions therefore arise similar to the one raised by ribosomal defects in MDS (particularly in relationship to the 5q- syndrome) - does an *SF3B1* mutation confer a clonal advantage to the mutant clone in the bone marrow compartment, and if so, how? What are the initiating mutational events? What is the origin of these mutations within the haematopoietic compartment? How *SF3B1* mutations may interact with

other coexisting spliceosome or known myeloid specific mutations? What is their prognostic significance in isolation or in combination with other mutations in MDS? Future analyses of *SF3B1* gene and known myeloid specific gene mutations in a larger diverse MDS patients will help to build a complete picture of the diseased genome and an understanding of the interactions and/or mutual exclusivities between genetic mutation(s), as well as clonal selection in MDS.

Chapter 4

SF3B1 and other spliceosome mutations exhibit specific associations with epigenetic modifiers and proto oncogenes mutated in MDS

4.1 Somatic mutations in MDS

Developments in genomics utilizing highly sensitive and scalable array-based and second-generation sequencing-based technologies have led to the identification of genes that are often mutated in MDS. Recent studies have reported mutations involving multiple components of the mRNA splicing machinery including *SRSF2, U2AF1*, *ZRSR2*, *PRPF40B*, *U2AF65* and *SF1* in patients with MDS, myeloproliferative disorder (MPN) and AML (Abu Kar et al., 2012; Damm et al., 2012a; Damm et al., 2011; Graubert et al., 2012; Makishima et al., 2012; Malcovati et al., 2011; Papaemmanuil et al., 2011; Patnaik et al., 2012; Thol et al., 2012b; Visconte et al., 2012a; Wu et al., 2012; Yoshida et al., 2011a). In addition, we and other have also reported mutations in splicing factor *SF3B1* gene in MDS patients with ring sideroblasts (Papaemmanuil et al., 2011; Smith et al., 2011; Yoshida et al., 2011b). *SF3B1*, the most frequently mutated spliceosome component in MDS (30% of cases), is mutated in 70-85% of RARS/RCMD-RS patients and correlate with the presence of ring sideroblasts (Papaemmanuil et al., 2011; Patnaik et al., 2012).

Over the past decade a number of novel gene mutations have been identified that are associated with MDS including genes involved in epigenetic regulation (*TET2, DNMT3A, IDH1/2, ASXL1, EZH2)* (Boultwood et al., 2010; Kosmider et al., 2010; Nikoloski et al., 2010; Smith et al., 2010; Walter et al., 2011) and cell signalling/transcription regulation (*TP53, NRAS/KRAS, RUNX1, FLT3, CCBL* and *ETV6)* (Bains et al., 2011; Bejar et al., 2011; Jadersten et al., 2011; Shih et al., 2004). In fact, around 50% of MDS patients have defects in one or more of these 'epigenetic' or 'oncogenic' factors, therefore suggesting an underlying genomic instability and aberrant transcriptional regulation in the evolution of this disease. However, establishing the relevance of these individual alterations with the pathogenesis of the disease has been a major challenge primarily because the molecular basis of MDS is phenotypically multifaceted and likely contributors are epigenetic changes, altered response to cytokines, the immune response to evolving clone(s), and the bone marrow stroma.

To gain a better understanding of *SF3B1* and other spliceosome aberrations, and how they may interact with other known myeloid specific gene mutations, and to determine their prognostic significance in isolation or in combination, a comprehensive mutational

screening was performed in a cohort of 154 MDS patients of various WHO subtypes (Table 29). Sequential samples were available from two patients in whom disease transformation (i.e. AML) was observed.

Patients with MDS seen at King's College Hospital from June 2004 to June 2014 were enrolled in this study. All patients had provided written informed consent in accordance to National Research Ethics Protocol (08/H0906/94). Demographic and clinical characteristics of the studied patients are detailed in Table 29. All patients were risk stratified according to IPSS categories. The clinical variables, FAB, WHO subtype and the prognostic risk of all patients were ascertained at the time of sample collection. The median follow-up was 21.4 months (Range, 1-83 months). The patient cohort was followed up to January 2012 for disease progression, and survival. The survival data for patients who underwent allogeneic haematopoietic stem cell transplant (HSCT) (n=35, 22%) were censored on the day of transplant.

A panel of 22 genes consisting of splicing factor genes *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*; genes implicated in epigenetic regulation *TET2*, *IDH1/2*, *ASXL1*, *EZH2* and *DNMT3A*; and known oncogenes/genes involved in cell signalling/transcription regulation *TP53*, *FLT3*, *NRAS*, *KRAS*, *RUNX1*, *CCBL*, *C-KIT*, *JAK2*, *MPL*, *CEBPA*, *BRAF* and *NPM* was selected and optimized for the mutational screening. These genes were selected based on the previously published reports showing the commonly mutated genes in myeloid malignancies. PCR primers specific to coding exons of all genes were designed using Primer 3 program. These primers were tagged with a universal sequencing primer (aUS1 for forward; aUS2 for reverse) which was subsequently used alongside Roche multiple-identifiers (MIDs) to barcode each patient (see methods section 2.2.5 for more details).

DNA isolated from 154 patient bone marrow total nucleated cells (TNCs) or $CD34^+$ cells was used for gene amplicon library preparation. Constitutional DNA (skin, n=27; $CD34^-CD3^+$ cells, n=18) was available from 45 patients. Following the PCR amplification of the exonic regions of the target genes, all amplicons (exonic regions) for each patient were pooled in one reaction. PCR amplified gene amplicons (after second round PCR) for randomly selected patients were loaded on an agarose gel and visualized under the UV-illuminator. In addition, the concentration of all gene amplicons (second round PCR) for 5 randomly selected patients was measured by

using the Quant-iT pico-green dsDNA assay kit as described in methods section 3.2.2. Pooled libraries were further processed for sequencing in a batch-wise manner (see methods section 2.2.2 for more details).

The sequencing was performed on Roche/454 FLX Sequencer. The average sequencing coverage across all genes was 200X and >90% of the coding regions had a coverage of >100X. This coverage enabled a reliable detection of mutant clones down to ≥5-10% MAB (defined as the proportion of sequence reads containing the mutation). Mutations in any one or more of these genes were detected in 76% (n=117/154) of the patients, with 38% (n=59) of the patients having splicing factor mutations and 49% (n=75) of patients harbouring more than one mutation (Figure34 and appendix table 15). Known single nucleotide polymorphisms (SNPs) and insertion/deletion variants listed in national Center for Biotechnology informations SNP database (dbSNP, build 140) and previously reported as germ line were excluded from further analysis. All mutations were confirmed through independent PCR and GS-FLX sequencing/Sanger sequencing experiments. For data relating to samples at/prior to transformation to AML, where 454-amplicon sequencing data was not available, relative peak intensity from Sanger sequencing was used to estimate MAB. Novel mutations which had not been previously reported were confirmed as acquired by their absence in 48/54 cases for whom constitutional source of DNA was available: skin biopsy (n=27), $CD3^+$ T-cells (n=18) and buccal swab (n=3) paired constitutional DNA. The remaining 7 novel variants (3 stop codons, 1 frame-shift mutation, 1 in-frame-shift deletion and 2 splice-site mutation) detected in 6 patients were also included in the analysis although constitutional source of DNA was not available (Appendix Table 19). The nature of these mutations makes them unlikely to be benign inherited variants. The reminder of the mutations identified in this study had previously been reported in literature as acquired mutations.

Table 29: Clinical characteristics of patients studied.

Patients were stratified by presence or absence of splicing factor mutations *SF3B1*, *SRSF2* and *U2AF1*. Cytogenetics failed in 4 patients in this cohort. P values with statistical significance are highlighted in bold. n- Represents number of patients; % - Represents percentage of the patients.

Patient Characteristics	Overall	SF3B1 Mutant	SF3B1 Wild Type	P value	SRSF2 Mutant	SRSF2 Wild Type	P value	U2AF1 Mutant	U2AF1 Wild Type	P value
	154	24 (16%)	130 (84%)		20 (13%)	134 (87%)		15 (10%)	139 (90%)	
Age, years				0.42			0.16			0.8
Median	65.5	65.2	63.1		66.9	63.1		62.8	63.5	
Range	17-85	35-83	17-85		51-82	17-85		48-75	17-85	
Sex				0.98			0.5			0.38
Male [n (%)]	104 (67%)	16 (66%)	88 (68%)		12 (60%)	92 (69%)		12 (80%)	92 (66%)	
WHO category *				**<0.001**			**<0.001**			0.2
RA/RCMD [n (%)]	40 (26%)	0 (0%)	40 (31%)		6 (30%)	34 (25%)		4 (27%)	36 (26%)	
RARS/RCMD-RS [n (%)]	24 (16%)	20 (83%)	4 (3%)		0 (0%)	24 (18%)		0 (0%)	24 (17%)	
RAEB -1/2 [n (%)]	49 (32%)	1 (4%)	48 (37%)		7 (35%)	42 (32%)		5 (33%)	44 (32%)	
s AML [n (%)]	15 (10%)	2 (8%)	13 (10%)		0 (0%)	15 (11%)		4 (27%)	11 (8%)	
t MDS/AML [n (%)]	12 (8%)	0 (0%)	12 (9%)		0 (0%)	12 (9%)		1 (7%)	11 (8%)	
CMML/MPD/MDS-U [n (%)]	14 (9%)	1 (4%)	13 (10%)		7 (35%)	7 (5%)		1 (7%)	13 (9%)	
Bone marrow blasts				0.19			0.39			0.27
Median (%)	5	1	6		8.5	4		9	4	
Range	0-80	0-19	0-72		0-19	0-80		0-59	0-80	
IPSS cytogenetic risk group	151			**<0.001**			0.15			0.46
Good [n (%)]	90	22 (96%)	68 (53%)		14 (70%)	76 (58%)		7 (50%)	83 (61%)	
Intermediate [n (%)]	17	1 (4%)	16 (13%)		4 (20%)	13 (11%)		3 (21%)	14 (11%)	
Poor N (%)	44	0 (0%)	44 (34%)		2 (10%)	42 (31%)		4 (29%)	40 (28%)	
Transfusion dependency				**0.04**			0.3			0.19
Yes [n (%)]	80 (51%)	17 (71%)	63 (47%)		8 (40%)	72 (52%)		10 (66%)	70 (49%)	
Progression to AML				**0.02**			**0.02**			0.4
Yes [n (%)]	44 (28%)	2 (8%)	42 (31%)		10 (50%)	34 (24%)		3 (20%)	41 (29%)	

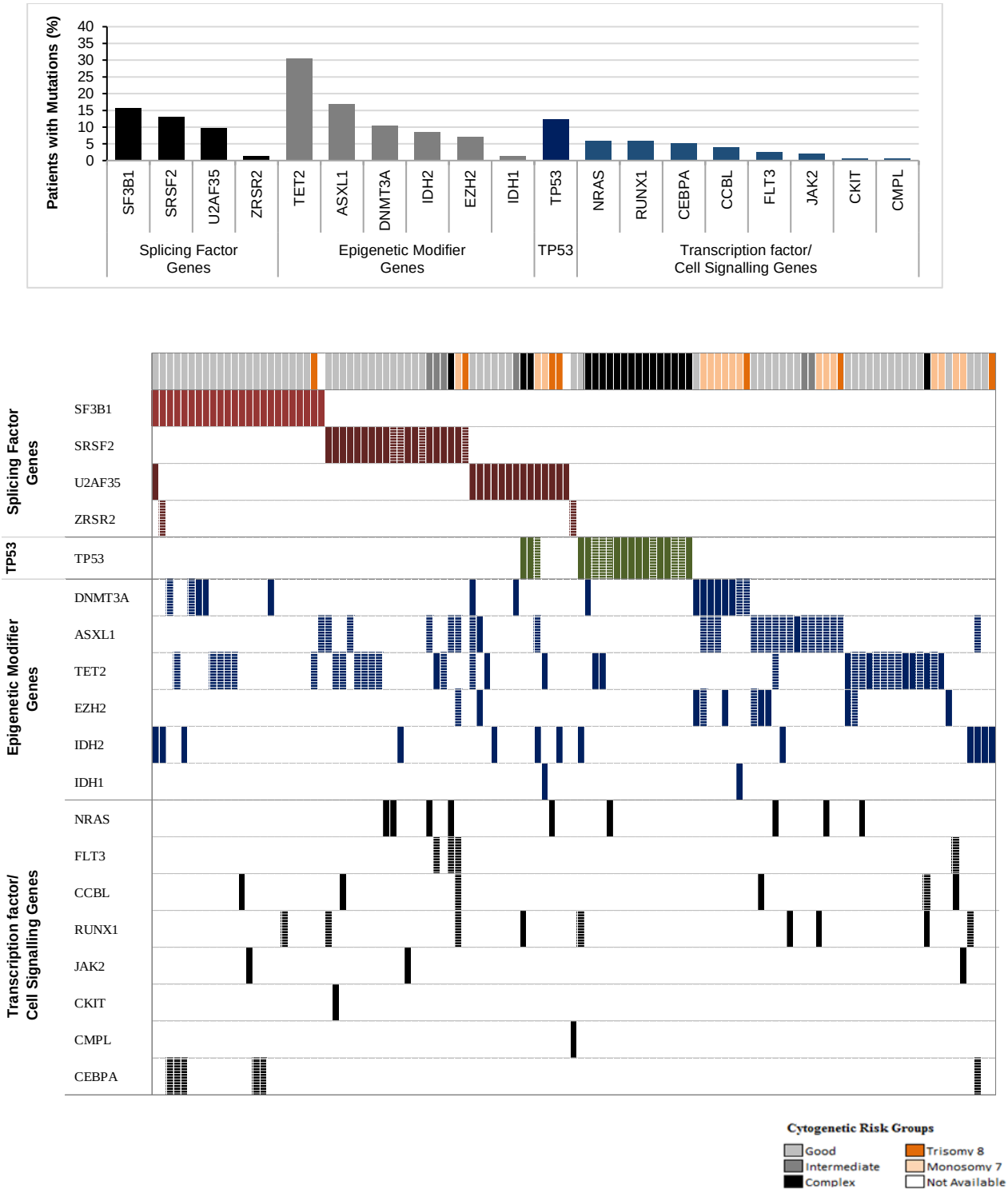


Figure 34: Gene Mutations detected in our MDS patient cohort.

(A) Frequency of gene mutations in our cohort of patients. (B)Top row represents distribution of mutations in 117 cases; where the shade of the bar indicates the cytogenetic risk groups according to the inset key. Underlying rows represent individual gene mutations denoted by coloured bars and specified on the left hand side. Bars with white strips indicate nonsense mutations, including splice-site mutations, while bars without stripes represent missense mutations.

4.1.1 Spliceosome gene mutations - frequency and clinical correlates

Sequencing of the splicing factor genes revealed 61 somatic mutations in 38% (59 of 154) of MDS patients (Table 29 and Appendix Table 19), comprising *SF3B1* 16% (n=24), *SRSF2* 13% (n=20), *U2AF1* 10% (n=15) and *ZRSR2* 1% (n=2). WHO subgroups for mutations of splicing factor genes included, RARS/RCMD-RS (20/24, 83%), CMML or MDS/MPN (9/14, 64%), sAML (6/15, 40%), RAEB-1/2 (13/49, 27%), RA/RCMD (10/40, 25%), but were uncommon in tMDS (1 of 12). Splicing factor mutations were common in low/int-1 IPSS (36/68, 53%) compared to int-2/high risk IPSS categories (12/63, 19%, p<0.002). Furthermore, with the exception of one case, all patients with isolated splicing gene mutations and no additional coexisting mutations (n=21) belonged to low/int-1 IPSS subgroup. Patients with complex karyotypes were less likely to harbour any splicing factor mutations as compared with good and intermediate risk IPSS cytogenetic groups (13 % *vs* 47% and 44%, p<0.002).

Overall, 24 of 154 (16%) MDS patients had a somatic mutation of *SF3B1*, however the frequency of *SF3B1* mutation was significantly higher in the RARS/RCMD-RS (20/24, 83%) as compared to other WHO categories (4/130, 3%, p<0.001). Although there were no significant differences in patients between the groups with respect to age, sex, blast percentage and neutrophil count, *SF3B1* mutations correlated strongly with lower haemoglobin (median-8.9 *vs* 10.1 g/dl, p<0.006), higher platelet count (median-296 *vs* 102 x 10^9/l, p<0.001), low/int-1 risk IPSS score (31% *vs* 0%, p<0.001), normal cytogenetics (24% *vs* 0%, p<0.002), transfusion dependency (21% *vs* 9%, p<0.03) and a decreased likelihood of leukaemic progression (4% *vs* 15%, p<0.02) when compared with wild type *SF3B1*.

Among the 154 MDS patients, 20 had *SRSF2* mutations (13%) and showed a significantly high neutrophil count (median-11 *vs* 2.8 x 10^9/l, p<0.001) and higher haemoglobin levels (median 10.9 *vs* 9.8 g/dl, p<0.03) compared to patients with wild type *SRSF2*. There was no difference in platelet count or transfusion dependency rate between the groups. *SRSF2* mutations were more frequently seen in patients with MDS/MPN or CMML (50%) and RAEB-1/2 (14%), but was absent in low risk IPSS including patients with ringed sideroblasts. Interestingly, both patients with isochromosome17q (n=2) had mutations of the *SRSF2* gene which maps to 17q25.1.

There was significant difference in rates of leukaemic progression in patients with mutant *SRSF2* compared with wild type (50% *vs* 24%, p<0.02), with two thirds of patients with co-existing *SRSF2* and cell signalling/transcription regulator mutations progressing to AML. *U2AF1* mutations were detected in 15 (10%) patients with clustering in male patients (12/15, 80%) and was also associated with lower haemoglobin (median-9 *vs* 10 g/dl, p<0.05) when compared to wild type *U2AF1*, but no significant differences in age, platelets, neutrophils, transfusion dependence, IPSS score or WHO subtype, were observed between the two groups. *ZRSF2* mutations were detected in only two (1%) MDS patients.

Interestingly, splicing factor mutations were largely mutually exclusive to each other, with only 2 patients having two separate spliceosome gene mutations, one with mutations in *SF3B1* and *U2AF1* whilst the other patient had mutations in *SF3B1* and *ZRSR2* genes.

4.1.2 Splicing factor mutations- Type, site and allele burden

Majority of *SRSF2* mutations were heterozygous non-synonymous amino acid substitutions, with an average MAB of 37.5% (n=20), comprising Pro95His/Leu/Arg (n=16) (Figure 35) changes as previously reported (Yoshida et al., 2011a). A novel 24 base pair deletion in *SRSF2* causing the frameshift mutation Tyr93fsX121 was detected in 4 patients, which would be predicted to cause loss of protein function (Figure 36).

However, *U2AF1* mutations exclusively present within the amino- and the carboxyl-terminal zinc finger motifs had a lower average MAB of 27% (n=15). These mutations targeted codon S34F (n=4) and Gln157Pro/Arg/His (n=11) of the protein (Figure 35). On the other hand, *ZRSR2* mutations had a much higher average MAB of 60.5% (n=2), were distinct in nature and did not cluster in the same protein domain (Figure 35), comprising in-frame deletion (Ser439_Arg440del) and a frame-shift deletion Glu133Glyfs11X.

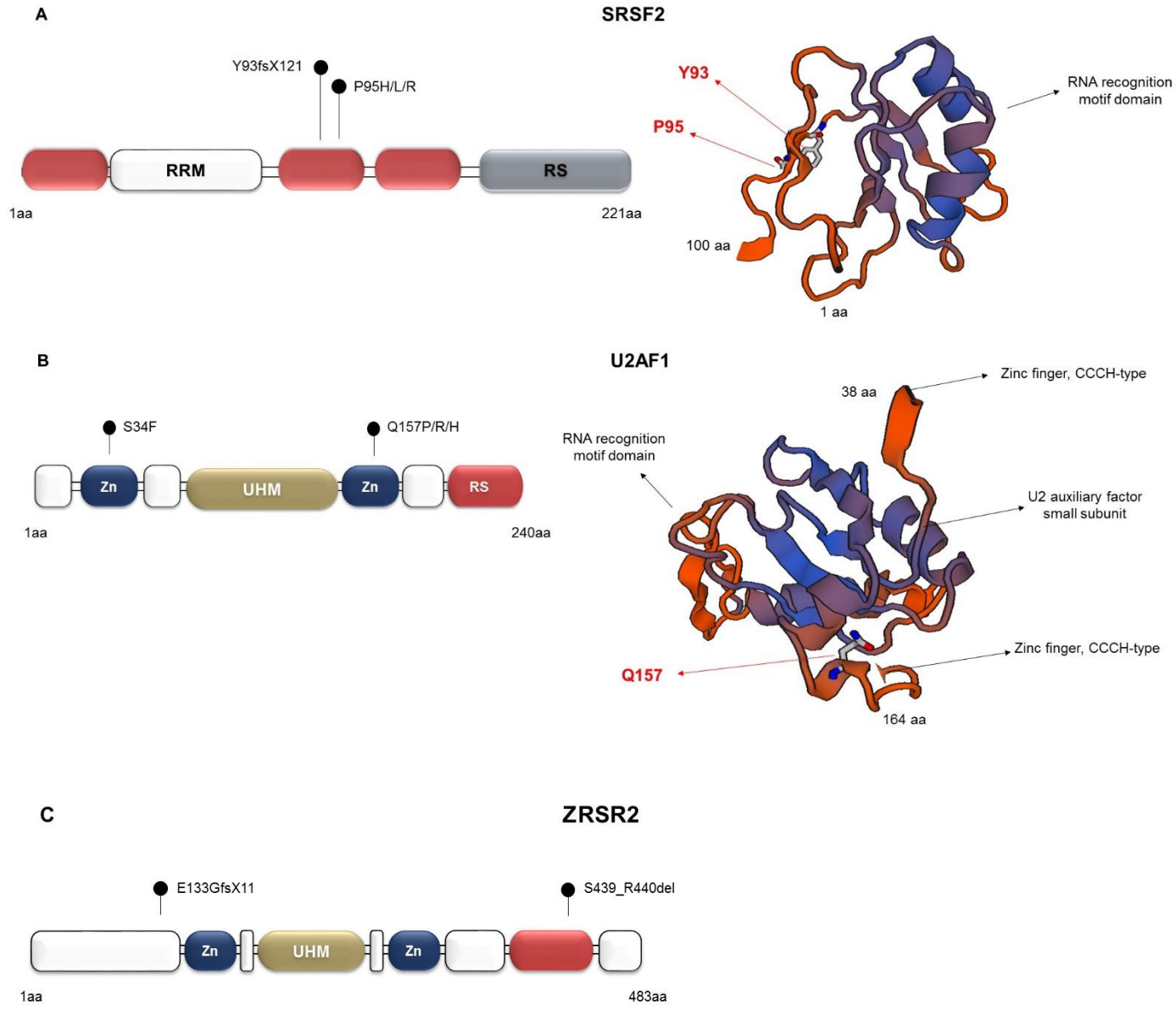


Figure 35: Distribution of *SRSF2, U2AF1* and *ZRSR2* mutations.

Each panel represents the known protein functional domains and 3D protein structure for *SRSF2, U2AF1* and *ZRSR2*. Distribution of somatic mutations are shown across all the protein domains for all genes.

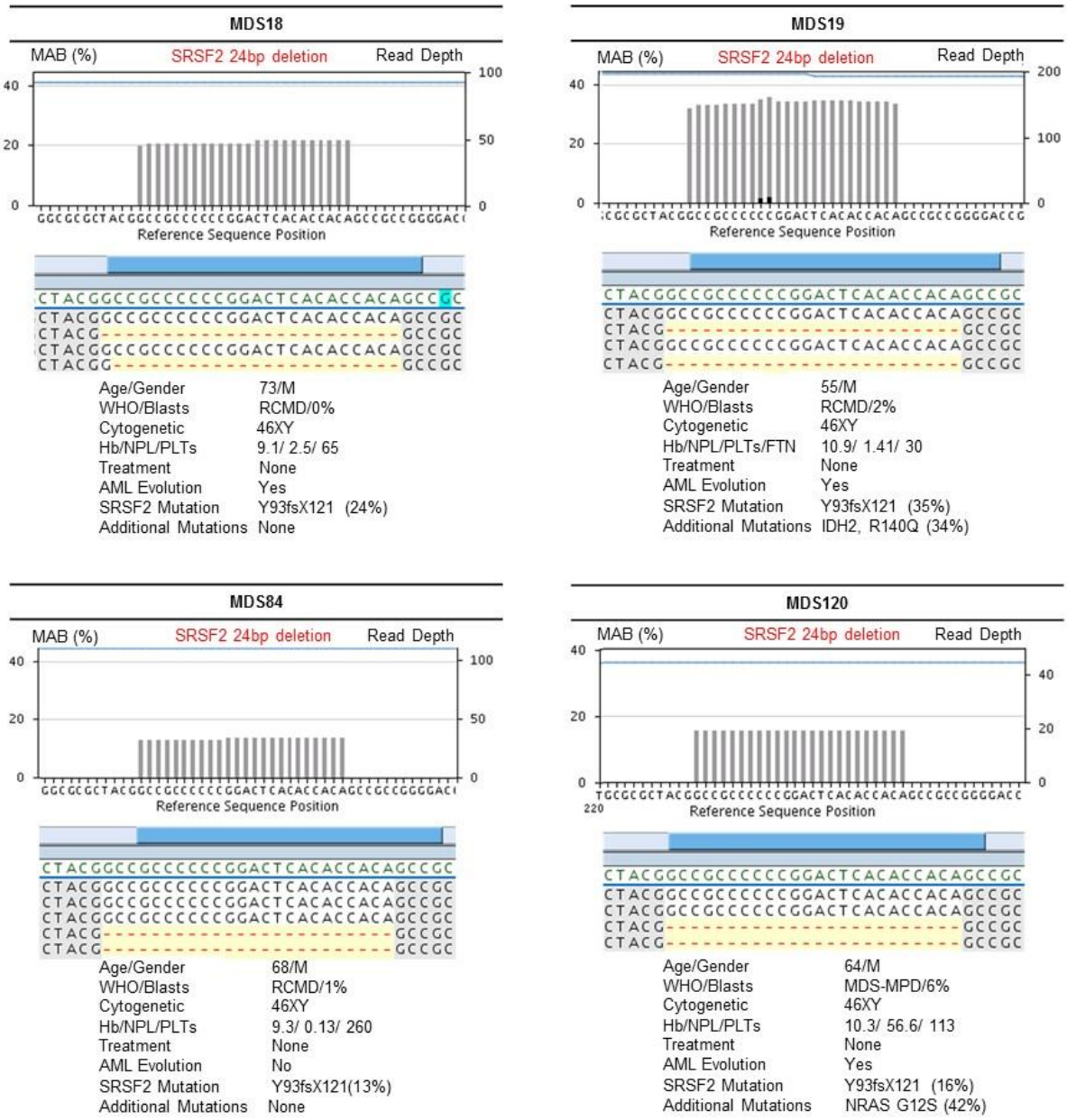


Figure 36: Patients with novel *SRSF2* mutations.

Top half of each panel shows the 454-Roche sequencing data output for patients with 24 basepair *SRSF2* deletion. Grey lines within the graph represents deleted nucleotides. Bottom half of each panel shows the clinical characteristics (Age, WHO, Blast count, Haemoglobin, Neutrophils, Platelets) as well as additional mutations (if present) for each patient. MAB-Mutant allele burden, Hb-Haemoglobin, NPL-Neutrophil count, PLTs-Platelets.

4.1.3 Prevalence of mutations in Epigenetic modifiers and Cell Signalling /Transcription regulators

The overall frequency of mutations of genes involved in epigenetic regulation (*TET2*, *ASXL1*, *DNMT3A*, *IDH2*, *EZH2* and *IDH1*), cell signalling/transcription regulators (*FLT3*, *RUNX1*, *NRAS*, *C-KIT*, *CCBL*, *JAK2* and *MPL*) and mutations in tumour suppressor gene *TP53* were 52% (n=80), 18% (n=28) and 12 % (n=19), respectively (Table 29 and Appendix table 19). Interestingly, mutations predicted to effect epigenetic regulation were detected in nearly half the patients, with *TET2* mutations being the most frequent in 22% (n=34) of the cohort. The frequency of other mutations was *ASXL1* (17%, n=26), *DNMT3A* (10%, n=16), *IDH2* (8%, n=13), *EZH2* (7%, n=11) and *IDH1* (1%, n=2). Although mutations in epigenetic modifiers clustered in female patients (64% *vs* 25%, p<0.05), there was no differences in the WHO subtypes, IPSS score, transfusion dependency or leukemic transformation rate when compared with cases wild type for genes involved in epigenetic regulation. Furthermore, mutations in genes involved in cell signalling/transcription regulation were detected in 18% of MDS patients with mutations of *NRAS* and *RUNX1* each present in 6% (n=9 each), *CCBL* (4%, n=6), *FLT3* (3%, n=4) and *JAK2* (2%, n=3). *C-KIT* and *MPL* mutations were detected in one patient each. Overall patients with these mutations, with the exception of *JAK2* and *MPL*, were associated with high risk MDS, increased blast count, transfusion dependency and increased likelihood of leukemic transformation when compared to their wild type counterparts (p<0.01). No *BRAF*, *NPM* and *KRAS* mutations were detected.

4.1.4 Mutual exclusivity of TP53 with spliceosome mutations

TP53 mutations were detected in 12% (19/154) of cases (Table 29 and Appendix table 19). *TP53* mutations were infrequent in patients with splicing factor mutations (5%, 3/59) compared to patients with wild type splicing factor genes (17%, 16/99, p<0.04). However, among patients with splicing factor mutations, all three *TP53* mutations were observed exclusively in those who had mutations in spliceosome component *U2AF1* (20% *vs* 0%, p<0.01). Among the patients with concurrent *TP53* and *U2AF1* mutations, two had complex cytogenetics and one had trisomy 8 (to be discussed further in discussion chapter).

4.1.5 Coexistence and exclusivity of splicing factor mutations with commonly mutated epigenetic modifiers and Cell signalling/transcription regulating genes

Of the 59 patients with spliceosome mutations, 15 (25%) had isolated splicing factor mutations, whilst 29 (49%) and 15 (25%) had mutations in epigenetic modifiers and cell signalling/transcription regulator mutations respectively, including 8 patients with coexisting mutations from all 3 mutation classes (Figure 34, Figure 37 and Table 30). Regardless of disease subtypes, MDS cases with non-*SF3B1* splicing factor mutations had significantly more mutations in other genes screened here (mean 2.4 mutations/case), than in patients with *SF3B1* mutations (mean 1.9 mutation/ case, $p<0.03$).

Furthermore, mutations of epigenetic modifiers were associated with mutant *SRSF2* compared to wild type (70% *vs* 50%, $p<0.07$), which was predominantly due to the presence of more *TET2* mutations with mutant *SRSF2* than wild type (40% *vs* 25%, $p<0.04$). Patients with splicing factor mutations were less likely to have multiple epigenetic modifier mutations (14%, n=5) compared to patients with wild type splicing factors (33%, n=15) ($p<0.03$), although if present, multiple mutations of epigenetic modifiers more often coexisted with *U2AF1* (n=4) rather than *SF3B1* or *SRSF2* (n=1) mutations. *DNMT3A* mutations were less likely to be seen with *SRSF2* mutations compared with other spliceosome mutations (0% *vs* 15%, $p<0.08$), whilst *ASXL1* mutations were more likely to co-occur with *SF3B1* than other splicing factor mutations (23% *vs* 4%, $p<0.07$), indicating non-random mutation associations and tolerances. A trend towards an association between *IDH2* and *U2AF1* was observed, when compared with other splicing mutations ($p<0.06$).

Mutations of genes involved in cell signalling/transcription regulation (*FLT3*, *NRAS*, *CCBL*, *RUNX1*, *JAK2*, *MPL* and *C-KIT*) clustered with splicing factor mutations (27% *vs* 13%, $P<0.02$). When splicing factor mutations were examined individually, these non-*TP53* mutations co-existed frequently with *SRSF2* mutations (50% *vs* 13%, $p<0.003$), compared to MDS patients with wild-type *SRSF2*. Furthermore, when considering all 59 cases of splicing factor mutants alone, mutations of *FLT3*, *NRAS*, *CCBL*, *RUNX1*, *JAK2*, *MPL*, *C-KIT* and *TP53* were significantly less often associated with *SF3B1* mutations compared with other splicing factor alteration cases (8% *vs*

34%, p<0.009). *SRSF2* mutations co-existed with mutations of cell signalling/transcription regulation genes, especially with alterations of *NRAS* (p<0.04) and *FLT3* (p<0.03), compared with non-*SRSF2* splicing factor mutations (40% *vs* 13%, p<0.02). Furthermore, *NRAS* mutations (n=9) were mutually exclusive to epigenetic modifier mutations *IDH2* (n=13) and *EZH2* (n=11). *CEBPA* mutations co-existed with *SF3B1* mutations than with other splicing factor mutations (21% *vs* 0%, p<0.008).

The average MAB for *SF3B1* and *SRSF2* mutations was 41% (24 cases) and 37.5% (20 cases), respectively. Coexisting point mutations present alongside with these splicing factor mutations had an average MAB of 35% and 37.5%, respectively, where the MAB of epigenetic modifiers (38% and 39%) was higher than cell signalling/transcription regulator mutations (27.5% and 28.75) for *SF3B1* and *SRSF2,* respectively. *U2AF1* mutations had a lower average allele burden of 27% (15 cases) with 39% average MAB of other coexisting mutations where the burden of the cell signalling/transcription regulator mutations (45%) was higher than epigenetic modifier mutations (33%). This seemed to be due to the *U2AF1*/*TP53* mutant cases which had a relatively higher *TP53* than *U2AF1* MAB in 2/3 cases. *ZRSR2* mutation allele burden was 60.5% (2 cases) on average.

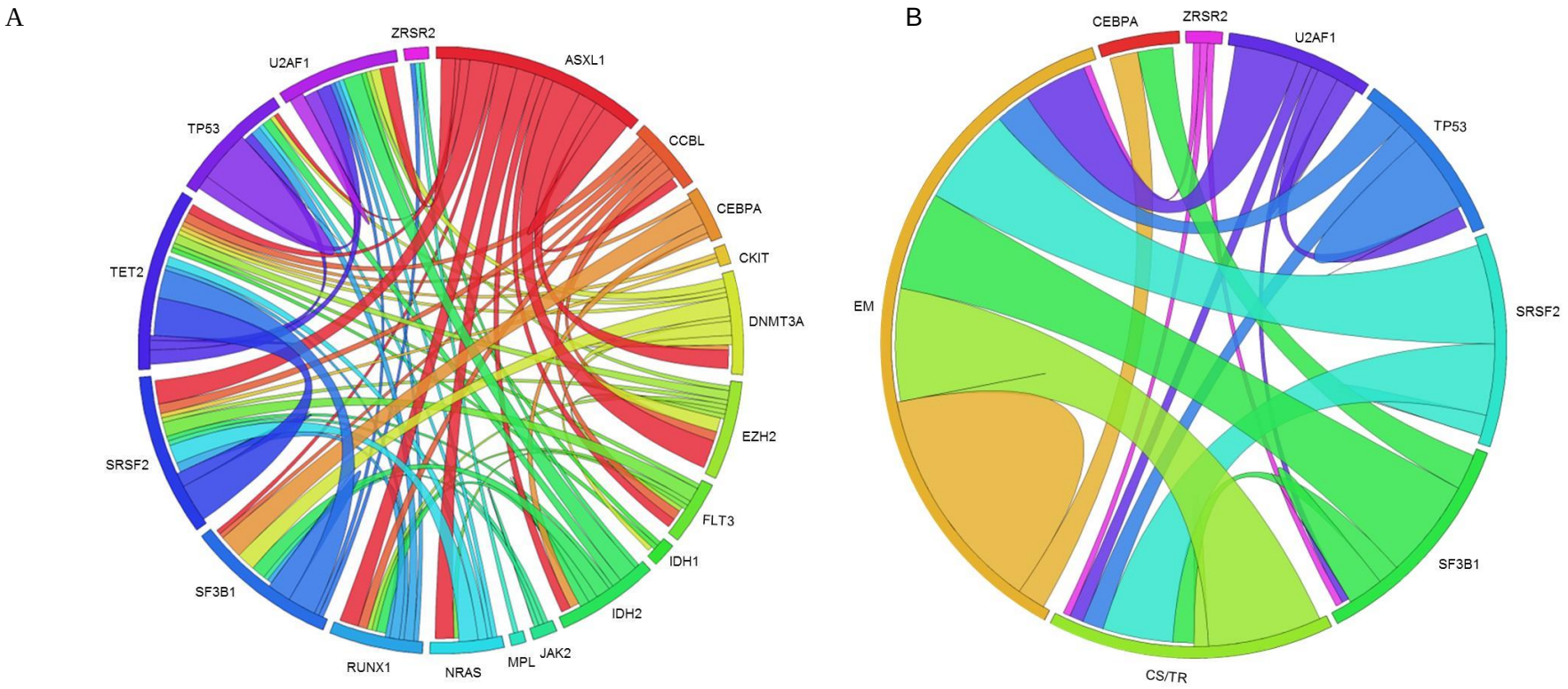


Figure 37: Frequency and distribution of various gene mutations in MDS patients as depicted by Circos diagram.

The arc length corresponds to the frequency of mutations in relevant gene and the ribbon width corresponds to the percentage of the patients that have other coexisting mutations. (A) Coexistent mutations between 22 individual genes in 119 patients. (B) Coexistent mutations between splicing factor genes (*SF3B1, SRSF2, U2AF1, ZRSR2*), epigenetic modifiers (EM), cell signalling/transcription regulation (CS/TF), *TP53* and *CEBPA*.

Table 30: Summary of mutations coexisting with mutant and wildtype spliceosome components in 154 patients.

Mutations were grouped according to their functional relevance; splicing factor, *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*; epigenetic modifier, *TET2*, *IDH1/2*, *ASXL1*, *EZH2*, and *DNMT3A*; Cell signalling/transcription regulator, *FLT3*, *NRAS*, *C-KIT*, *RUNX1*, *CCBL*, *JAK2* and *MPL*; tumour suppressor *TP53*. Table 2A compares mutations in the grouped genes. Table 2B shows co-existing gene mutations and individual gene mutations coexisting with splicing factor mutations. 37 patients were wildtype for all screened genes. Table cells indicate number of patients within mutant or non-mutant spliceosome groups followed by the percentage where appropriate.

A	Overall	SF3B1 Mutant	SF3B1 WT	P value	SRSF2 Mutant	SRSF2 WT	P value	U2AF1 Mutant	U2AF1 WT	P value
Patients (n, %)	154	24 (16%)	130 (84%)		20 (13%)	134 (87%)		15 (10%)	139 (90%)	
CO-EXISTING GENE MUTATIONS										
Epigenetic Modifier Mutations (n, %)	80 (52%)	14 (58%)	66 (51%)	0.5	14 (70%)	66 (49%)	0.07	9 (60%)	71 (54%)	0.6
Cell Signalling/ Transcription regulator mutations (n, %)	28 (18%)	3 (12%)	25 (19%)	0.5	10 (50%)	18 (13%)	0.03	2 (13%)	26 (19%)	0.9
TP53 Mutations (n, %)	19 (12%)	0 (0%)	19 (15%)	0.04	0 (0%)	19 (14%)	0.07	3 (20%)	16 (12%)	0.4

B	**Mutant Spliceosome**	**SF3B1**	**SRSF2**	**U2AF1**	**ZRSR2**	**Wild Type Spliceosome With Other Mutations**
Total mutant cases	**59**	**24**	**20**	**15**	**2**	**58**
Cases with additional co-existing non-splice factor mutations	**48** (81%)	**19** (79%)	**17** (85%)	**12** (80%)	**2** (100%)	**27** (47%)
Cases with >1 additional co-Existing non-splice factor mutations	**16** (27%)	**4** (17%)	**7** (35%)	**5** (33%)	**0**	**5** (9%)
CO-EXISTING GENE MUTATIONS						
TP53	**3** (5%)	**0**	**0**	**3** (20%)	**0**	**16** (28%)
FLT3	**3** (5%)	**0**	**3** (15%)	**0**	**0**	**1** (2%)
NRAS	**5** (8%)	**0**	**4** (20%)	**1** (7%)	**0**	**4** (7%)
RUNX1	**4** (7%)	**1** (4%)	**2** (10%)	**1** (7%)	**0**	**5** (9%)
CCBL	**3** (5%)	**1** (4%)	**2** (10%)	**0**	**0**	**3** (5%)
C-KIT	**1** (2%)	**0**	**1** (5%)	**0**	**0**	**0**
JAK2	**2** (3%)	**1** (4%)	**1** (5%)	**0**	**0**	**1** (2%)
MPL	**1** (2%)	**0**	**0**	**0**	**1** (50%)	**0**
ASXL1	**9** (15%)	**1** (4%)	**5** (25%)	**3** (20%)	**0**	**17** (29%)
DNMT3A	**7** (10%)	**5** (20%)	**0**	**2** (13%)	**0**	**9** (16%)
EZH2	**2** (3%)	**0**	**1** (5%)	**1** (7%)	**0**	**9** (16%)
TET2	**17** (29%)	**6** (25%)	**8** (40%)	**3** (20%)	**0**	**17** (30%)
IDH2	**7** (12%)	**3** (13%)	**1** (5%)	**4** (27%)	**1** (50%)	**6** (11%)
IDH1	**1** (2%)	**0**	**0**	**1** (7%)	**0**	**1** (2%)
CEBPA	**5** (8%)	**5** (21%)	**0**	**0**	**0**	**1** (2%)

4.1.6 Sequential acquisition of cell signalling/transcription regulating gene mutations in *SF3B1* mutant *clones* with disease transformation

Only 2 of 24 patients with RARS/RCMD-RS and *SF3B1* mutations evolved to AML at variable time points after diagnosis (Patient MDS54, 4 months; patient UPN MDS105, 36 months). To investigate the contributions of *SF3B1* and coexisting mutations in disease evolution, relevant sequential samples were screened for both these cases. Sequencing analysis of the samples taken from patient MDS54 and MDS105 at the time of diagnosis and at presentation to our institute (24 months after diagnosis) revealed *SF3B1* mutation with a MAB of 39% and 42%, respectively (See exampled for MDS54 in Figure 38). A *TET2* mutation was detected at diagnosis, with a MAB of 60%, in MDS54 (Figure 38) and a *RUNX1* mutation in MDS105. However, MDS105

transformed to AML and underwent allogeneic HSCT in morphological remission, following induction chemotherapy. He relapsed with AML and lost his donor chimerism, shortly after HSCT. The *SF3B1* and *RUNX1* mutation burden was maintained at the same level as at the RARS stage and at AML relapse after HSCT in MDS105. Similarly, *SF3B1* and *TET2* mutations were maintained at the same mutant allele burden at transformation in MDS54. At the time of transformation to AML, MDS54 also acquired a mutation in *RUNX1* (Phe163Tyr) with a MAB of ≈30%, and a *FLT3-ITD* (Phe590_Trp603dupInsPro) with a MAB of ≈50% (Figure 38). Following intensive chemotherapy, MDS54 attained a transient morphological remission but relapsed rapidly with *FLT3-ITD* and *RUNX1* MAB increasing to ≈80% and ≈45% respectively. The MAB of *SF3B1* and *TET2* genes remained constant at 40% to 50% throughout the disease duration, which as a heterozygous mutation would occur in a majority of sample cells. As such, *RUNX1* and *FLT3* mutations seemed to have evolved from the *SF3B1*/*TET2* clonal population and coexist in the same cells (Figure 38).

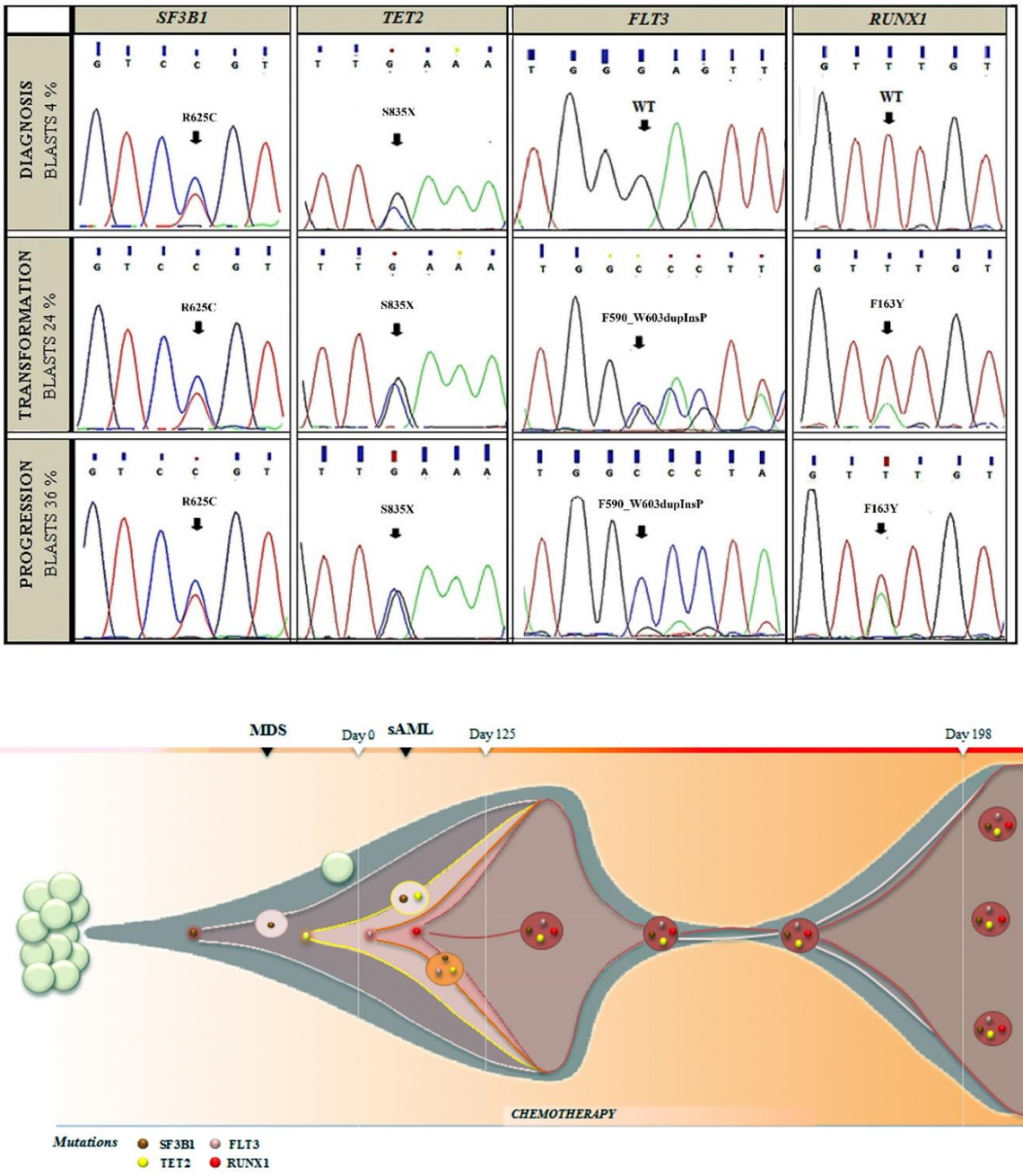


Figure 38: Clonal evolution and disease progression in RARS patient.

Sequencing analysis of sequential samples in MDS54 with an *SF3B1* mutation. Three horizontal rows represent samples collected at different time points: diagnosis, transformation to AML and disease progression after a short duration of remission, respectively. Accumulation of the oncogenic mutations and changes in mutant allele burden levels are seen through disease progression along with increase in the blast count. Sanger sequencing was used to confirm/ determine the mutation status throughout the experiment. 454 sequencing confirmed the mutation level differences at start and end points of the experiment. WT-Wildtype

4.1.7 Prognostic significance of mutations

The median overall survival (OS) of the entire cohort was 34.4 months (95% confidence interval (CI), 16.9 to 51.9 months). Univariate analysis revealed that patients with *SF3B1* mutations had a better OS [Not Reached (NR) *vs* 24.2 months, p<0.003] and progression free survival (PFS) (NR *vs* 40.3, p<0.02) than wild type *SF3B1* (Figure 39A and Figure 39B), whilst none of the other splicing factor gene mutations had an impact on either outcome measure. The median OS of patients with either of splicing factor mutations was significantly better than wild type spliceosome (NR *vs* 24.2 months, p<0.03), but no difference in PFS was seen between the groups (Figure 39C and Figure 39D).

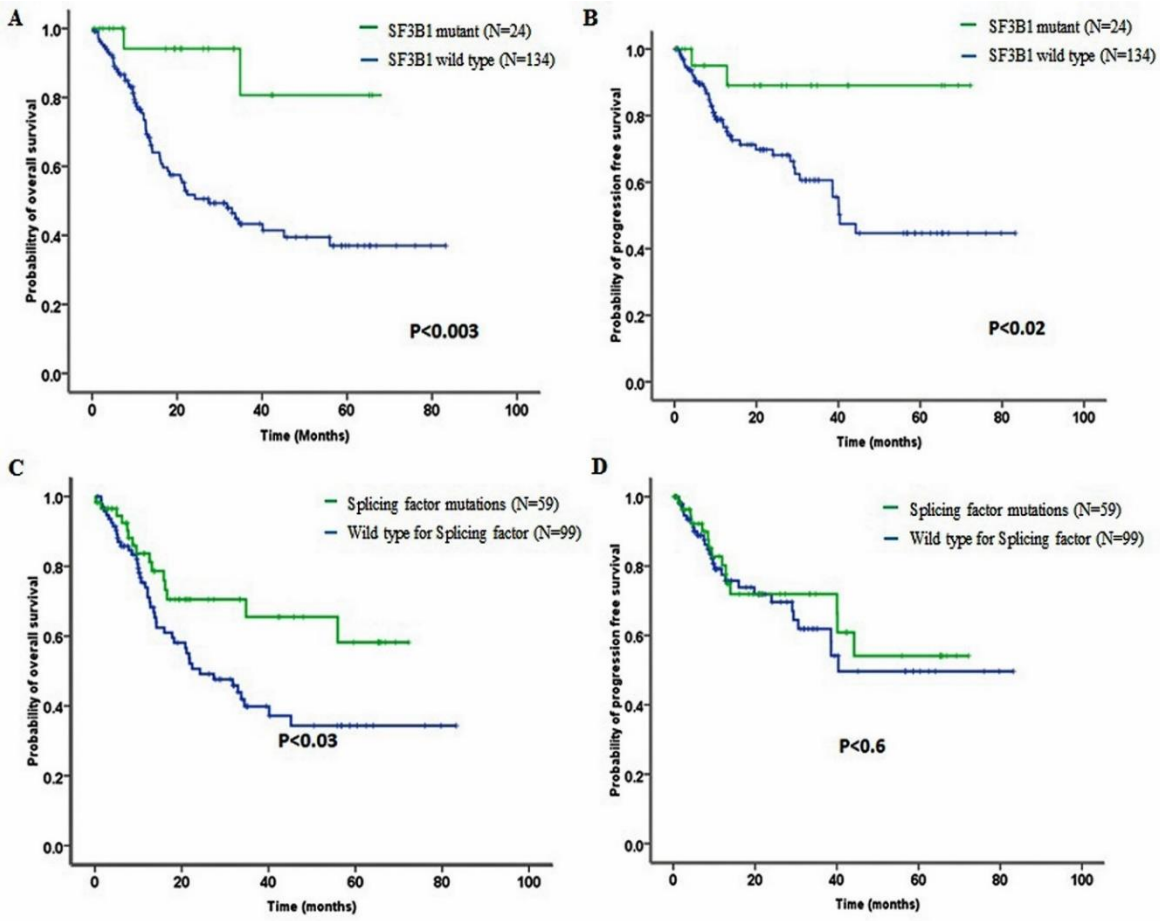


Figure 39: Statistical analysis for *SF3B1* and Splicing factor gene mutations.

Overall survival (A) and progression free survival (B) for patients with SF3B1 mutations (n=24) compared with wild type SF3B1 (n=130). Overall survival (C) and progression free survival (D) for patients with spliceosome mutations (n=59) compared with patients without splicing factor mutations (n=99).

On a univariate model, there was no significant difference in either OS or PFS in patients with other individual mutations compared to wild type, except for *TP53* and *NRAS* (Figure 40). Among patients with epigenetic modifier mutations (n=81), a proportion also had coexisting splicing factor mutations (n= 37) and 44 patients had epigenetic modifier mutations alone. A trend towards a better survival was seen in the group of patients with epigenetic modifier mutations with splicing gene mutations compared with the rest (p<0.06), although the distribution of mutations of genes involved in cell signaling /transcription regulation and *TP53* mutations were evenly distributed in both groups (Figure 41).

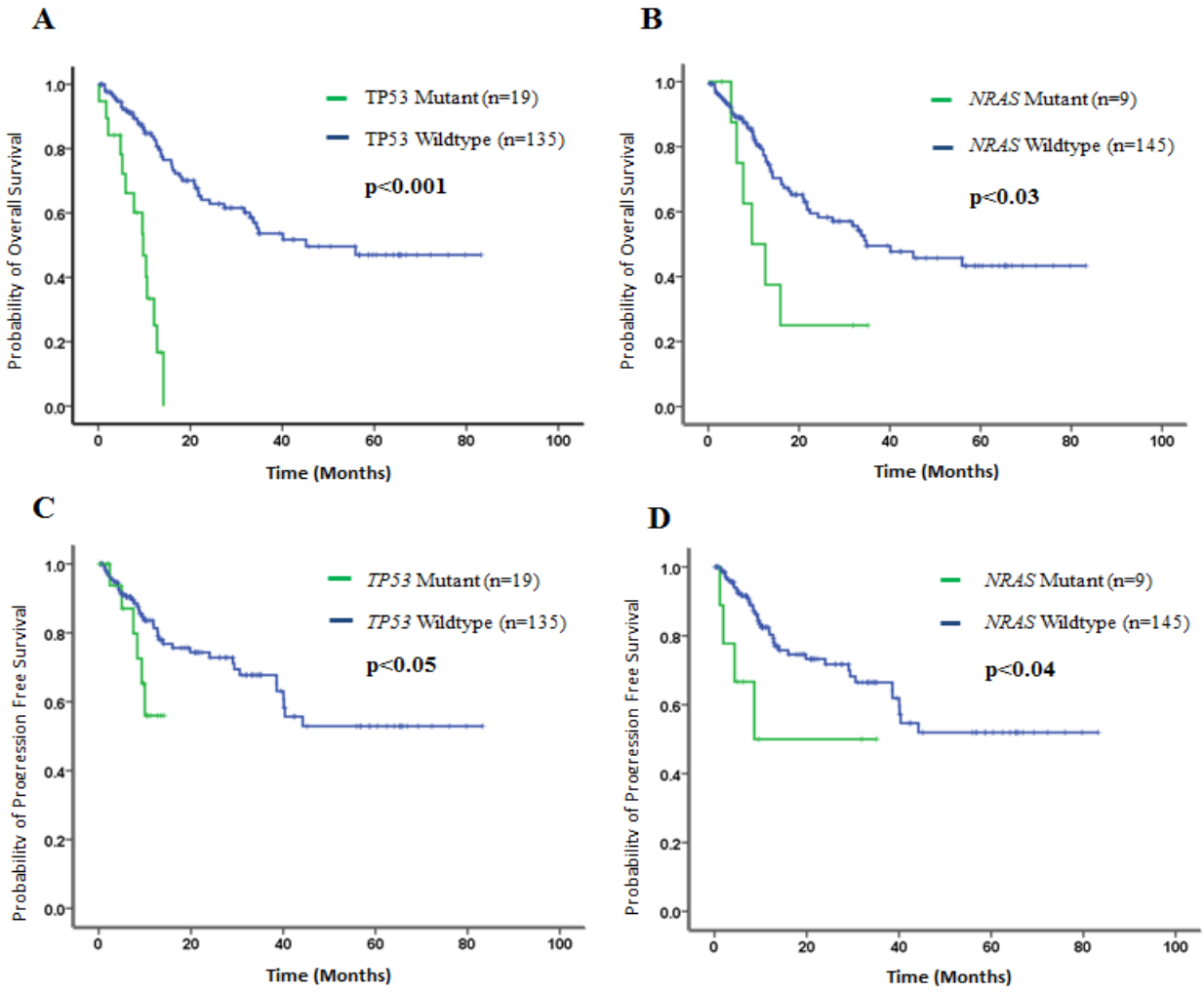


Figure 40: Statistical analysis for *TP53* and *NRAS* gene mutations.

Overall survival (A) and progression free survival (C) for patients with *TP53* mutations (n=19) compared with wild type *TP53* (n=135). Overall survival (B) and progression free survival (D) for patients with *NRAS* (n=9) compared with patients without *NRAS* mutations (n=145).

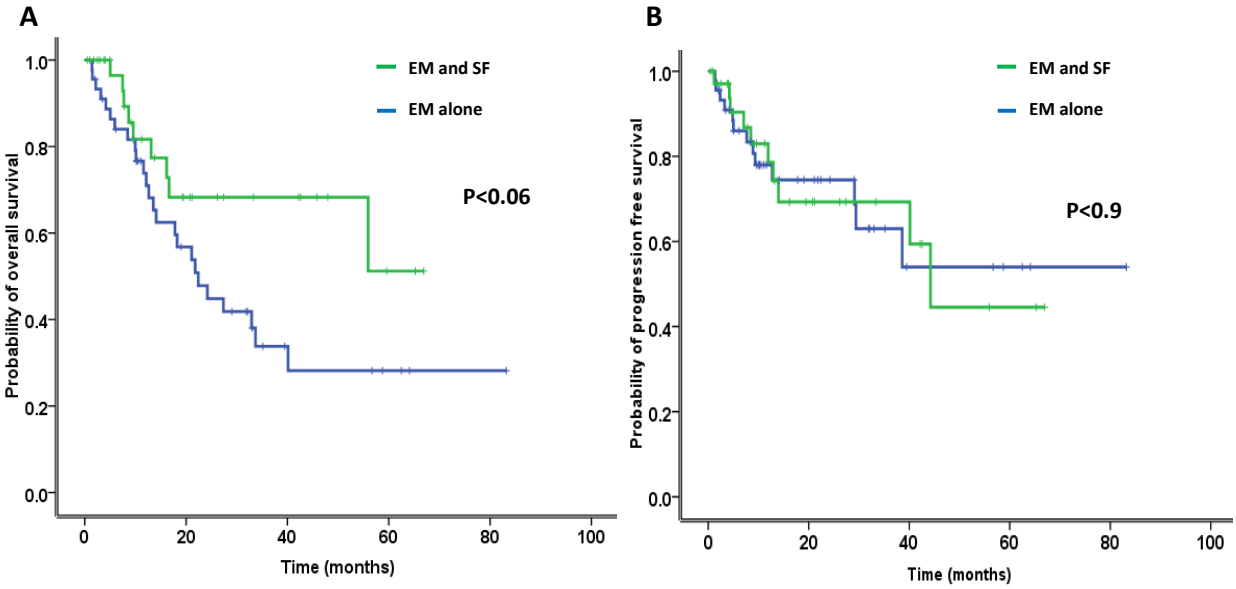


Figure 41: Statistical analysis for Epigenetic modifying and splicing factor gene mutations.

Overall survival (A) and progression free survival (B) for patients with epigenetic modifier (EM) mutations stratified according to coexistence of splicing factor (SF) mutations with epigenetic modifier mutations (n=37) or epigenetic modifier mutations alone (n=41).

Patients with mutations of splicing factor and of genes involved in cell signalling/transcription regulation or *TP53* (n=15) had an extremely poor OS (15.8 months *vs* NR, p<0.009) and PFS (12.5 months *vs* NR, P<0.001) when compared with spliceosome mutations without mutations of genes involved in cell signalling/transcription regulation (n=44), (Figure 42) although this group had only 3 patients with *TP53* mutation.

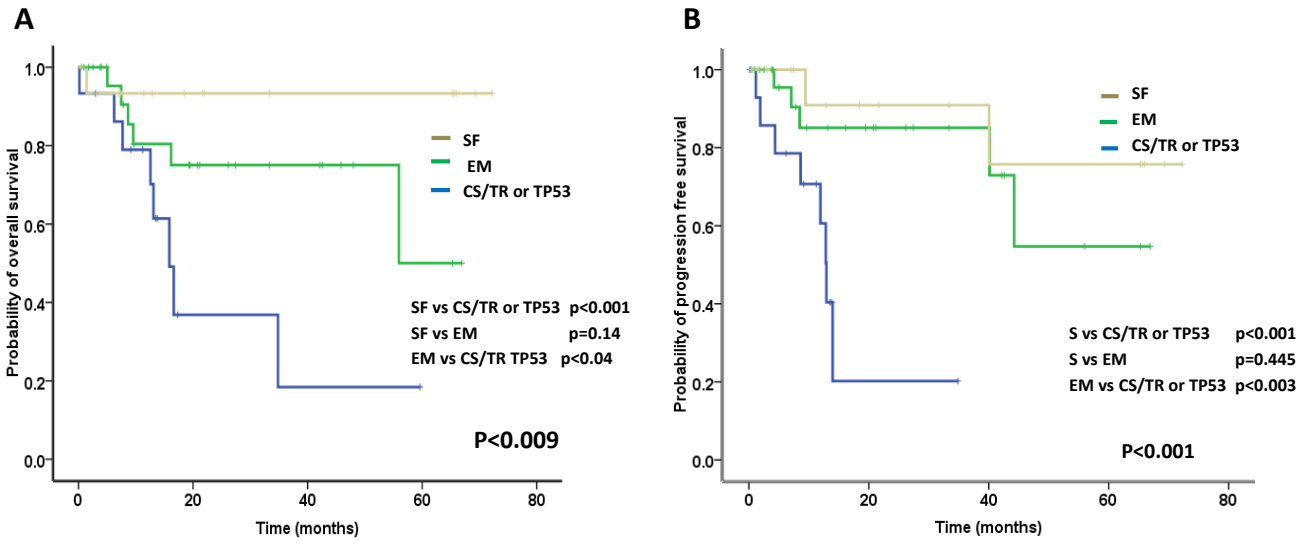


Figure 42: Overall survival (A) and progression free survival (B) for patients with spliceosome mutations (n=59).

Patients were stratified according to patients with splicing factor mutations (SF) co-existing with mutations of genes involved in cell signalling/transcription regulation (CS/TR) or *TP53* mutations versus splicing factor mutations co-existing with epigenetic modifier (EM) mutations versus patients with splicing factor mutations alone.

Table 31: Multivariable analysis of overall survival.

The variables included are age, WHO category, medullary blasts, IPSS cytogenetic risk groups, transfusion dependency, *SF3B1* mutations, *NRAS* mutations and *TP53* mutations.

Variable	Hazard Ratio	OS	
		95% CI	p value
Age	1.0	1-1.1	0.05
WHO category	1.1	0.9-1.3	0.12
IPSS Cytogenetic	1.9	1.3-2.6	0.001
Medullary blast count	1	0.9-1	0.2
Transfusion dependency	1.7	0.9-3	0.08
SF3B1	0.2	0.1-0.8	0.03
TP53	2.1	1.1-4.4	0.04

Multivariant analysis using variables: age at the sample collection, WHO category, medullary blast count, IPSS cytogenetic groups, transfusion dependency status, and *SF3B1*, *NRAS*, *TP53* mutations, revealed that *NRAS* mutations did not have impact on OS or PFS. Patients with *SF3B1* and *TP53* mutations had varying impacts on OS with hazard ratios (HR) of 0.2 ($p<0.03$, 95% CI, 0.1-0.8) and 2.1 ($p<0.04$, 95% CI, 1.1-4.4), respectively. None of the analyzed genotypes, including *SF3B1* and *TP53*, impacted on PFS in multivariable model (Table 31). The results this chapter will be further discussed in the discussion chapter.

4.2 Discussion

Our initial findings and also other concurrent reports (Goldstein et al., 2010; Makishima et al., 2012; Papaemmanuil et al., 2011; Visconte et al., 2011; Yoshida et al., 2011a) of frequent spliceosome pathway mutations in MDS in-general, prompted us to perform a comprehensive study of mutant spliceosomal genes including *SF3B1* in 154 MDS patients. To gain further understanding into the genetic and biochemical pathways driving this disease, and the interaction of the various known myeloid related gene mutations, we also investigated other genes commonly mutated in MDS. In this study, 76% of MDS patients had mutations of the genes screened, with nearly 50% of patients having more than one mutation. Two large sequencing studies using an 111gene panel (Papaemmanuil et al., 2013) and 104 gene panel (Haferlach et al., 2014) published subsequently reported a mutation rate of 78% and 89.5% in MDS patients, respectively. The mutation rate reported by Haferlach and colleagues is higher which can be due to the variability in the MDS subtypes between published studies including ours. For examples, 40% of the patients sequenced in the Haferlach study were RAEB1/RAEB11 patients which are known to have higher mutation frequency.

Splicing factor genes, *SF3B1, SRSF2*, *U2AF1* and *ZRSR2* were mutated in 38% of patients. Although collectively these mutations were found in a wide spectrum of MDS subtypes, some were strongly associated with specific disease features. *SF3B1* mutations for example were strongly correlated with high levels of ringed sideroblasts and were therefore found in 80% of RARS/RCMD-RS patients, as also reported in other studies (Haferlach et al., 2014; Papaemmanuil et al., 2011; Papaemmanuil et al., 2013). Conversely, *SRSF2* and *U2AF1* mutations were often seen in advanced forms of MDS such as RAEB1/RAEB11 and CMML respectively, which fitted well with a higher number of coexistent mutations in genes involved in cell signalling/transcription regulation (e.g. *NRAS*, *FLT3* and *RUNX1*) with known oncogenic functions.

Occurrence of *SF3B1* mutations has been linked to significantly better OS and in some instances longer, leukaemia-free and event-free survival in RARS (Papaemmanuil et al., 2011). Similarly a beneficial independent prognostic impact for *SF3B1* mutations

on outcome, especially OS was seen in this study. In accordance with this observation is an absence of coexistent *TP53* aberrations, which is a strong predicator of poor OS, according to data presented both here and elsewhere (Bejar et al., 2011). Furthermore, data from a cohort of 317 MDS patients indicated no influence of *SF3B1* mutations on OS or time to leukemic progression (Damm et al., 2012c). In addition to this, previous studies have also linked mutations of U2AF1 (Damm et al., 2012b; Makishima et al., 2012) and *SRSF2* (Makishima et al., 2012) with an increased risk to AML progression and/or shorter OS. However, another study by *Thol et al* demonstrated only *SRSF2* mutations were associated with shorter OS as well as time to AML progression and not mutations in *U2AF1 (Thol et al., 2012a)*. We could only demonstrate a correlation of *SRSF2* mutations with progression to AML, but no impact on OS. *U2AF1* mutations similarly did not have any impact on the outcome in this study. Although *NRAS* and *TP53* mutations had impact on outcome, only *TP53* retained significance in a multivariate model.

Within the spliceosome mutant group, *SF3B1* seemed to be the strongest driver of a beneficial effect and the only spliceosome factor independently attributed with better OS and PFS. Such a model possibly suggests that splicing factor mutations are early disease events and define a 'founder' disease clone that subsequently develops additional, increasingly deleterious mutations during the natural course of the disease, clonal evolution and disease progression. Furthermore, as particular splicing factor mutations differ in their patterns of association with other mutations, this suggests a hierarchy of tolerated mutational load and limitation to particular paths of disease evolution. For example, a dearth of co-existent *TP53* mutations throughout spliceosome mutant cases and mutual exclusivity of *TP53* mutations with *SF3B1* and *SRSF2* mutants at the time of diagnosis and disease transformation further supports a restrictive pattern of genetic insults. Furthermore, a restrictive pattern of additional coexisting mutations previously linked with disease transformation, such as *FLT3* and *RUNX1*, are seen alongside *SF3B1* mutations, although this might be expected in low risk disease such as RARS/RCMD-RS, which make up the majority of *SF3B1* mutants. However, it is important to note that *SF3B1* mutant cases which gained such oncogenic mutations during transformation maintained a constant *SF3B1* MAB, indicating evolution within the same disease clone population.

Interestingly, *SRSF2* mutants, which are present both in low and high risk disease cases, similarly did not coexist with *TP53* mutations here, but rather *FLT3*, *NRAS* or *RUNX1* mutant oncogenes, as was the case for transformed *SF3B1* mutants. Conversely, *U2AF1* mutations were found to coexist with *TP53* mutations here, although the *U2AF1* mutation burden was significantly lower than that of *TP53* in 2/3 of these cases, where *TP53* mutation burden was high and consistent with LOH on chromosome 17p confirmed by metaphase cytogenetics in one of these cases. Furthermore, *FLT3* and *NRAS* mutations were independently more likely to occur with *SRSF2* mutations, accentuating the leukemic transformation rate in these patients, but were less likely to occur with *SF3B1* mutant cases. The questions therefore remains, are such mutual exclusivities driven at a molecular level and due to a lethal combination of genetic lesions and do mutations in different spliceosome components carry different weights in terms of biological importance and different thresholds of additional insults that a particular mutant clone can endure?

Significantly, mutations of epigenetic modifiers also seemed to fall in with particular splicing factor mutations. For example, *ASXL1* mutation was less likely to coexist with *SF3B1* mutations compared to *SRSF2* and *U2AF1* mutations cases, whilst *DNMT3A* mutations were not found with *SRSF2* but coexisted with *SF3B1* mutants (4/24). Similar results were reported by Bejar and colleagues showing the molecular synergy between *SF3B1* and *DNMT3A* genetic lesions (Bejar et al., 2012). Again this highlights the importance of genetic background and how particular mutation combinations may attain an acceptable biochemical equilibrium and stable disease. It is of particular interest that only single epigenetic modifier mutations seemed to co-occur with *SF3B1* and *SRSF2* mutations, which would fit the theory that such mutations represent early events in disease where numerous clonal genetic lesions or separate disease clones have not yet evolved. This is again highlighted in the case of *TET2* mutations, reasoned as early disease events and which do not add a prognostic value in MDS (Delhommeau et al., 2009; Smith et al., 2010), which were more likely to be found in MDS cases with mutant spliceosome in our patient cohort, especially *SRSF2* mutations.

Mutation analyses performed on serial samples in patients who were initially diagnosed with *RARS* and subsequently transformed to AML, also provides us with

clear evidence that *SF3B1* mutations are an early ancestral event. In these cases, the *SF3B1* mutant clone survives various treatments during the course of the disease, acquiring additional oncogenic mutations such as *FLT3* and/or *RUNX1*, enabling the disease to evolve. Supporting this theory, splicing factor mutants in isolation have a better OS in MDS cases generally, which seemingly worsens as additional mutant genes are added to the genetic makeup. Bejar and colleagues in a previous study reported that mutations in 5 genes (*TP53*, *RUNX1*, *EZH2*, *ASXL1* and *ETV6*) are independent predictors of poor OS in patients with MDS (Bejar et al., 2011) which tends to support our data. It is worth to note that splicing factor gene mutations were not studied in this study. Therefore, the contribution of *SF3B1* mutations in disease transformation or progression to AML may be limited, but they may instead provide a favourable environment or sufficient pressure for other more destabilizing mutations to occur. Furthermore, the fact that *SF3B1* mutations are present in $CD34^{+}$ but not in T-cells in MDS, further reinforces their importance and link to a clonal advantage through the course of myeloid disease, particularly in RARS.

Tumour heterogeneity observed in the bone marrow is quite common in MDS, especially in advanced WHO subtypes and high-risk prognostic subgroups. This also seems to be associated with increasing number of particular gene mutations and intratumoral subpopulations (low level mutations) observed in these patients as reported here and elsewhere (Haferlach et al., 2014; Papaemmanuil et al., 2013). Taken together, accumulation of increasing number of gene mutations especially 'oncogenic' mutations and growing intra-tumoral heterogeneity characterize the clonal evolution as well as advancement of the MDS disease which is associated with a worse prognosis.

It is noteworthy that functional studies to date have not as yet arrived at a consensus in identifying definitive functional pathways affected by splicing factor mutations, even in RARS where *SF3B1* mutation dominate. The variation between RARS samples with *SF3B1* mutations may either be a function of the different *SF3B1* mutations, or reflect additional mutations present in the MDS clone in these cases. Although, the majority of the *SF3B1* mutant cases have a dominant heterozygous mutation clone, however, the percentage of ring sideroblasts in these patients ranges considerably from 20% to 95%. Interestingly, various studies (Cazzola et al., 2012; Damm et al., 2012c;

Papaemmanuil et al., 2011) including ours have found a relationship between the *SF3B1* MAB and the percentage of ring sideroblasts, however, this does not explain the huge phenotypic (i.e. ring sideroblasts percentage) variability in these patients. Therefore, this again implies the interplay of several factors driving the MDS clone, where spliceosome mutations are perhaps fuelling subtle biochemical changes which can be built upon in the course of disease evolution by the emergence of other changes (i.e. genetic or epigenetic) within the diseased genome.

Recent findings including ours have shown that multiple mutations are present in clonally evolving diseases such as MDS, brings forth a series of important questions: What are the initiating mutational events? What is the origin of these mutations within the haematopoietic compartment? Which of the mutation subsets are “driver mutations” and hence contribute to the tumour progression? Spliceosome mutations are largely mutually exclusively to each other and are less likely to have multiple epigenetic modifier mutations as compared to cases with wild-type spliceosome. Does this infer that mutations of multiple genes that have different functions or participate in independent pathways may be additive or even synergistic in providing an advantage to the tumour clone as compared to the mutations of two genes participating in the same pathway or process? Is there a “cross talk” between the processes or pathways? The relationship between the founding mutant clone to the “leukaemia initiating cell” and also the other emerging low-level mutant clones is not yet clear; therefore separation of individual diseased clonal populations and functional testing is essential to determine this partnership. Hence, identifying the combinatorial patterns of “driver mutations” in individual diseased clones would help us to identify the functional relationships between the mutated genes and aberrant pathways in MDS as well as predict their malignant potential.

Chapter 5

MDS initiating-cells originate from the haematopoietic stem cell compartment in *SF3B1* mutant acquired sideroblastic anaemia

5.1 Introduction

5.1.1 Haematopoiesis and Xenotransplantation

Since the pioneering work of Till and McCulloch in the early 1960s, the science of bone marrow-derived haematopoietic stem cells (HSCs) has progressed enormously from observational to functional studies. The regenerative potential of HSCs was further established with clonal *in-vivo* repopulation assays in two independent studies, thus confirming the existence of multipotential HSCs and their progeny (Becker et al., 1963; Till and Mc, 1961). This finding prompted others to develop clonal *in-vitro* assays that have been combined with recent developments of genomic technologies, stem cell isolation as well as xenotransplantation models in various studies. This has started to lead to a better understanding of the complex clonal architecture and mutational hierarchy of phenotypically and functionally defined 'malignant stem cells' detected in haematological malignancies. Furthermore, the successful establishment of murine xenograft NSG models, particularly for acute myeloid leukaemia (AML) has yielded empirical evidence for the existence of so-called 'cancer stem cells', a minor subpopulation of cells responsible for maintenance of neoplastic proliferation (Horton and Huntly, 2012). In addition, recent studies have demonstrated that chemotherapy-resistant leukemic stem cells reside in the endosteal region of bone marrow (Ishikawa et al., 2007). The findings from these studies have helped to clarify how leukemic cells are maintained and propagated *in-vivo*.

5.1.2 Haematopoietic stem cells

HSCs are undifferentiated multipotent stem cells with the ability to perpetuate themselves for indefinite periods through self-renewal. HSCs reside as rare cells (1 in 10^6 bone marrow cells) in the bone marrow.

To be defined as a "true stem cell", a cell must demonstrate durable self-renewal capacity as well as differentiation potential into all the downstream cell lineages within the haematopoietic hierarchy (Figure 43). It should also be do so at a clonal or single-cell level to exclude the possibility that a population that is homogeneous in terms of cell surface antigen expression is still functionally heterogeneous and composed of multiple single-lineage precursors. Functional *in-vitro* clonogenic assays (long term culture-initiating cells, LTC-ICs and CFC) as well as quantitative *in-vivo* transplantation

methods when combined with the FACS, has provided increasingly detailed insight into the stem cell biology of both normal and malignant haematopoiesis.

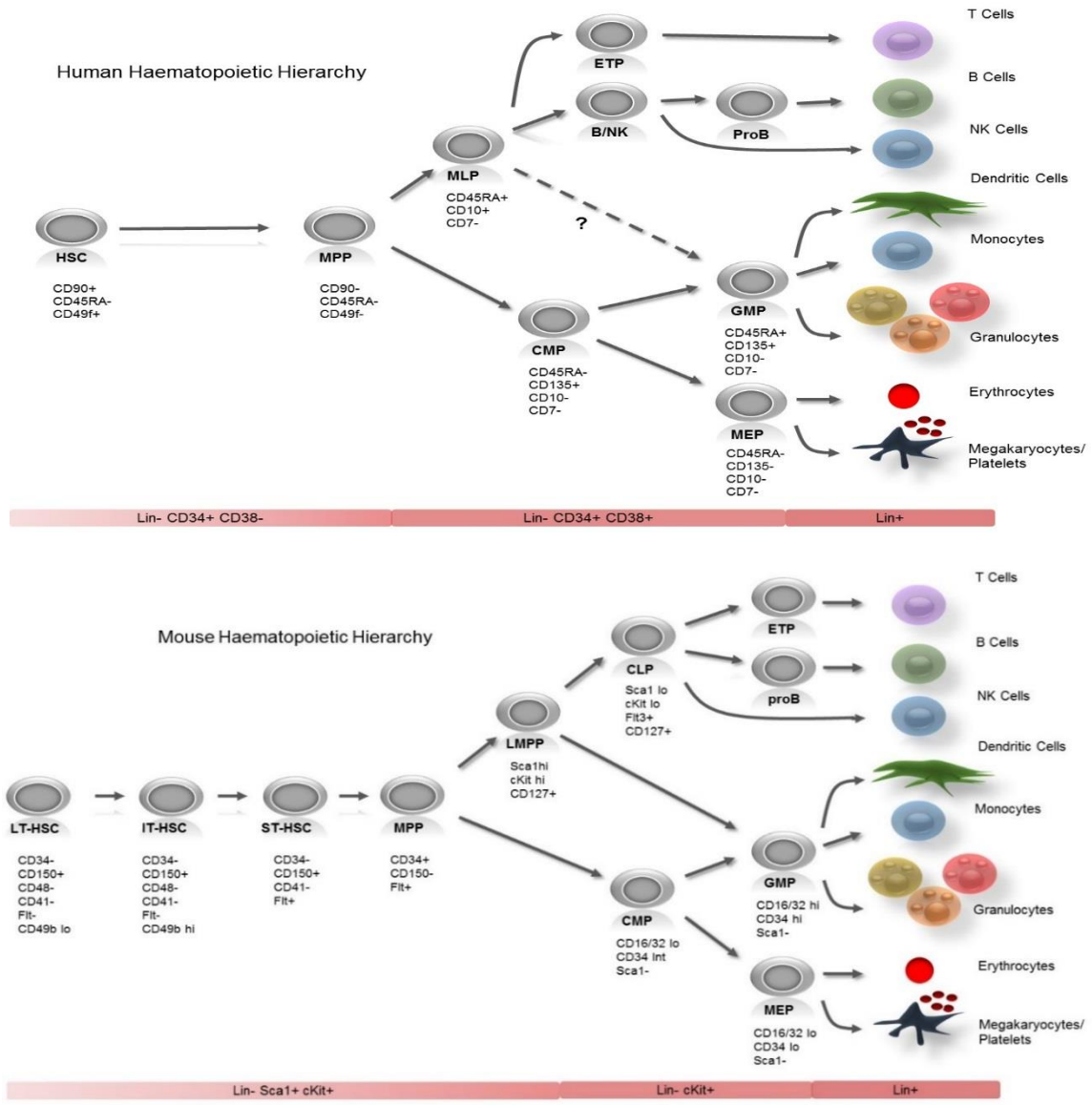


Figure 43: Current models of hematopoietic differentiation in humans (top) and mouse (bottom).

In humans (top), haematopoietic stem cells (HSCs) defined by CD49f marker have long-term reconstituting and self-renewing capacity; multipotent progenitors (MPPs) have limited self-renewal leading to transient but multilineage reconstitution; common myeloid progenitors (CMPs); common lymphoid progenitors (CLPs). In mice (bottom), HSCs can be divided into long-term (LT), intermediate-term (IT), and short-term (ST) classes based on the duration of repopulation capacity. In mice, differentiation of HSCs gives rise to transiently engrafting MPPs, and a series of immature lymphoid-biased progenitors (such as LMPPs) that undergo gradual lymphoid specification. Both human and mice have well-defined populations of myelo-erythroid progenitors: CMPs, GMPs and MEPs. CLPs are further differentiated into early T-progenitors (ETPs) and pro-B cells.

Over the past few decades, our ability to prospectively characterize mouse HSCs has improved but our understanding of the human HSCs lags far behind, especially in haematological malignancies. Like in the mouse, characterization and purification of human HSCs requires simultaneous detection of several independent cell specific markers. Cell surface antigen CD34, a ligand for L-selectin that is expressed on upto 3% of the human bone marrow cells was the first antigen found to enrich human HSCs as well as progenitors (Civin et al., 1984). Subsequent clinical studies have confirmed that it is a HSC specific marker (Kang et al., 2008; Vogel et al., 2000). *In-vivo* studies later demonstrated that the most primitive human HSCs express CD34 but lack CD38 cell surface antigens (Bhatia et al., 1997; Cashman et al., 1997; Hogan et al., 2002). However, cells expressing $CD90^+$ (thy1) within the $CD34^+CD38^-$ population have been shown to have the highest capacity to reconstitute the bone marrow of a sublethally irradiated NOD/SCID host (Baum et al., 1992; Murray et al., 1995). In addition, the human trials demonstrating the HSC function was performed using autologous mobilized peripheral blood in clinical transplantation, where long-term engraftment was observed by transplantation of purified $CD34^+CD38^-CD90^+$ cells (Michallet et al., 2000; Negrin et al., 2000; Vose et al., 2001). Further studies introduced CD45RA as marker of more differentiated progenitors that negatively enrich for HSCs (Majeti et al., 2007). Together, these studies support the idea that human HSCs are present within the $CD34^+CD38^-CD90^+CD45RA^-$ fraction of the haematopoietic compartment and they constitutes <0.5% of the total adult bone marrow cells.

Immediately downstream of HSCs are MPP that are not capable of long-term self-renewal. Until recently, little was known about these cells; however, the distinction between MPPs and HSCs is based on mouse studies. These studies have showed that a number of intermediate cell types with varying degrees of self-renewal capacity exist between long-term HSCs and the first lineage-committed progenitor. The first study to provide evidence for MPPs in humans came from an *in-vivo* experiment where transplantation of $CD34^+CD38^{lo}$ cord blood cells generated transient myelo-erythroid engraftment (Mazurier et al., 2003). Subsequently, Majeti and colleagues proposed that the loss of CD90 expression by $CD34^+CD38^-CD45RA^-$ cells distinguishes $CD90^+$ HSCs from transiently-engrafting $CD90^-$ MPPs. However, cells lacking CD90 antigen still engrafted after serial transplantation in NOD/SCID mice, suggesting that this

population was not completely resolved from HSCs (Majeti et al., 2007). *Notta et al* in a recent study reported that CD49f (integrin alpha 6) is also expressed on ≈50% of human $CD90^{+}$ and ≈25% of $CD90^{-}$ cells. In the same study, cells expressing CD49f marker within $CD34^{+}CD38^{-}CD45RA^{-}CD90$ cell population was shown to efficiently generate long-term multilineage grafts *in-vivo* and the loss of CD49f marker identified transiently engrafting MPPs (Notta et al., 2011).

The hierarchical progression from HSCs to mature progenitors within the bone marrow compartment involves gradual loss of self-renewal ability, proliferation capacity, and lineage potentials. However, the loss of self-renewal capacity by progenitors during differentiation precedes lineage commitment. Furthermore, the earliest commitment decision (downstream of MPPs) segregates myelo-lymphoid lineages which give rise to common myeloid progenitors (CMPs) and lymphoid progenitors, and each of these undergo further commitment steps. This also forms the basis of the first comprehensive “classical” model of the haematopoiesis in defining committed mouse myeloid and lymphoid lineages (Akashi et al., 2000; Kondo et al., 1997; Reya et al., 2001). The mouse mature progenitors have an extensive proliferative potential and generate large numbers of mature cells. More recently, efforts at defining human myeloid progenitor cell populations by cell surface markers have been attempted using data from mouse studies. Human myeloid progenitors (CMPs, GMPs, and MEPs) were isolated using IL-3 receptor a chain (CD123) or FLT3 (CD135), and CD45RA cell surface antigens. GMPs, but not MEPs, express CD123 and CD135 antigens (Doulatov et al., 2010). In addition, the human CMP→GMP transition is marked by acquisition of CD45RA antigen (an isoform of CD45 that negatively regulates some select classes of cytokine signalling), and $CD135^{+}CD45RA^{-}$ CMPs produced myeloerythroid cells, but not lymphoid compartment *in-vitro* as well as *in-vivo (Manz et al., 2002)*. However, the main question over the years has been whether CMPs always give rise to GMPs and MEPs, or whether there are other possible branch points that give rise to MEPs—for example, from an HSC or MPP. Using lineage tracing data with FLK2-Cre mice, it has been shown that the majority of mouse erythrocytes and megakaryocytes are derived from Flk2-positive multipotent progenitors. In mice, Flk2 is expressed by CMPs, but not in adult HSCs, MPPs and MEPs, therefore it seems that most of the MEPs and GMPs transition through a CMP stage of the differentiation.

Thus, human myeloid cell development seems to be consistent with the mouse haematopoietic classical model.

Unlike the myeloid progenitors CMPs/GMPs/MEPs, which follow the classical model of differentiation, lymphoid progenitors have been more controversial with some *in-vivo* studies suggesting an alternative "myeloid-based" model. According to this model, lymphoid and myeloid progenitor fates remain coupled, instead of splitting early in differentiation process (Kawamoto et al., 2010). This model predicts the existence of progenitors with myelo-lymphoid, but not E-MK, potential. In mouse, this hypothesis was validated with the isolation of LSK $CD34^+$ $Flt3^{hi}$ lymphoid-primed multipotent progenitors (LMPPs) which display priming of lymphoid transcripts and mediate transient lympho-myeloid repopulation with lymphoid bias, but have a very low E-MK potential. Therefore, the available evidence suggests that lymphoid differentiation is not a single lineage bifurcation, but perhaps a gradual and parallel differentiation process with many intermediate states. As a result of the uncertainties in determining the precise potential for any given progenitor population, particularly for human bone marrow progenitors, studies have proposed a broader term, namely MLP (Doulatov et al., 2010). MLPs are used to describe any progenitor that can give rise to all the lymphoid lineages (B-, T-, and NK- cells), but that may or may not have other lineage potentials (for example, GMPs).

The search for the cell surface antigens present on the early lymphoid progenitors MLPs in human bone marrow has been focused on the earliest T cell marker i.e. CD7 antigen and B cell marker i.e. CD10. Study by *Galy et al* demonstrated that $CD34^+CD10^+$ cells derived from human bone marrow were able to give rise to B-, T- and Nk cells, but not myelo-erythoid cells (Galy et al., 1995). However, Hao and colleagues used *in-vitro* single cell assay to show that rare $CD7^+$ cells present within the primitive $CD34^+CD38^-$ population were found to differentiate to B and NK cells, but not myelo-erythoid cells (Hao et al., 2001). Likewise, another recent study also reported robust T cell potential of this progenitor population in mice (Hoebeke et al., 2007). In a recent study, single-cell assay revealed that the multilymphoid (B, T, and NK) potential in humans was restricted to the $CD34^+CD38^-CD90^{-/lo}CD45RA^+$ ($CD90^-$ $CD45RA^+$) compartment. Human myeloid cells predominantly monocytes, macrophages, and dendritic cells were also present in the lymphoid colonies

generated from single CD90^{-}CD45RA^{+} cells. In addition, human CD90^{-}CD45RA^{+} cells transiently engrafted in NSG mice and gave rise to both myelomonocytic and B cells, arguing that their myeloid potential is not an artefact of *in vitro* assay (Doulatov et al., 2010). Another study reported similar observation using human CD90^{-}CD45RA^{+} cells. MLPs gave rise to B, T, and NK cells, as well as myeloid lineages *in-vitro* as well as in *in-vivo* assays. Engrafted human myeloid cells consisted of granulocytes and monocytes, suggesting these human MLPs have lineage potential similar to that of mouse LMPPs (Goardon et al., 2011). Therefore, it is clear that human MLPs are not lymphoid restricted, and also possess myeloid, but not erythro-myeloid potential.

5.1.3 NSG Xenotransplantation

The laboratory mouse (*Mus musculus*) has been one of the best model systems for elucidating the cancer biology. Out of the ≈23,000 protein coding genes known to be present in mouse genome, only 75% of genes are in 1:1 orthologous relationships with human genes (total reported protein coding genes ≈ 21,000). Despite such differences, mouse models have proven to be useful in validating gene functions, identifying novel cancer genes and tumour biomarkers as well as providing better clinical models in which to test novel therapeutic strategies. This is due to various benefits including its relatively small size, lifespan of 3 years, an entirely sequenced genome and molecular as well as physiological similarities to humans.

Over the past decade advances in the genetic technologies have led to the development of various mice models which can be broadly divided into two groups. First, genetically engineered transgenic and mutant mice that recapitulate a specific disease genotype; second group, the syngeneic and xenogeneic mice in which tumour cells are transplanted into immunocompetent or immunodeficient mice, respectively.

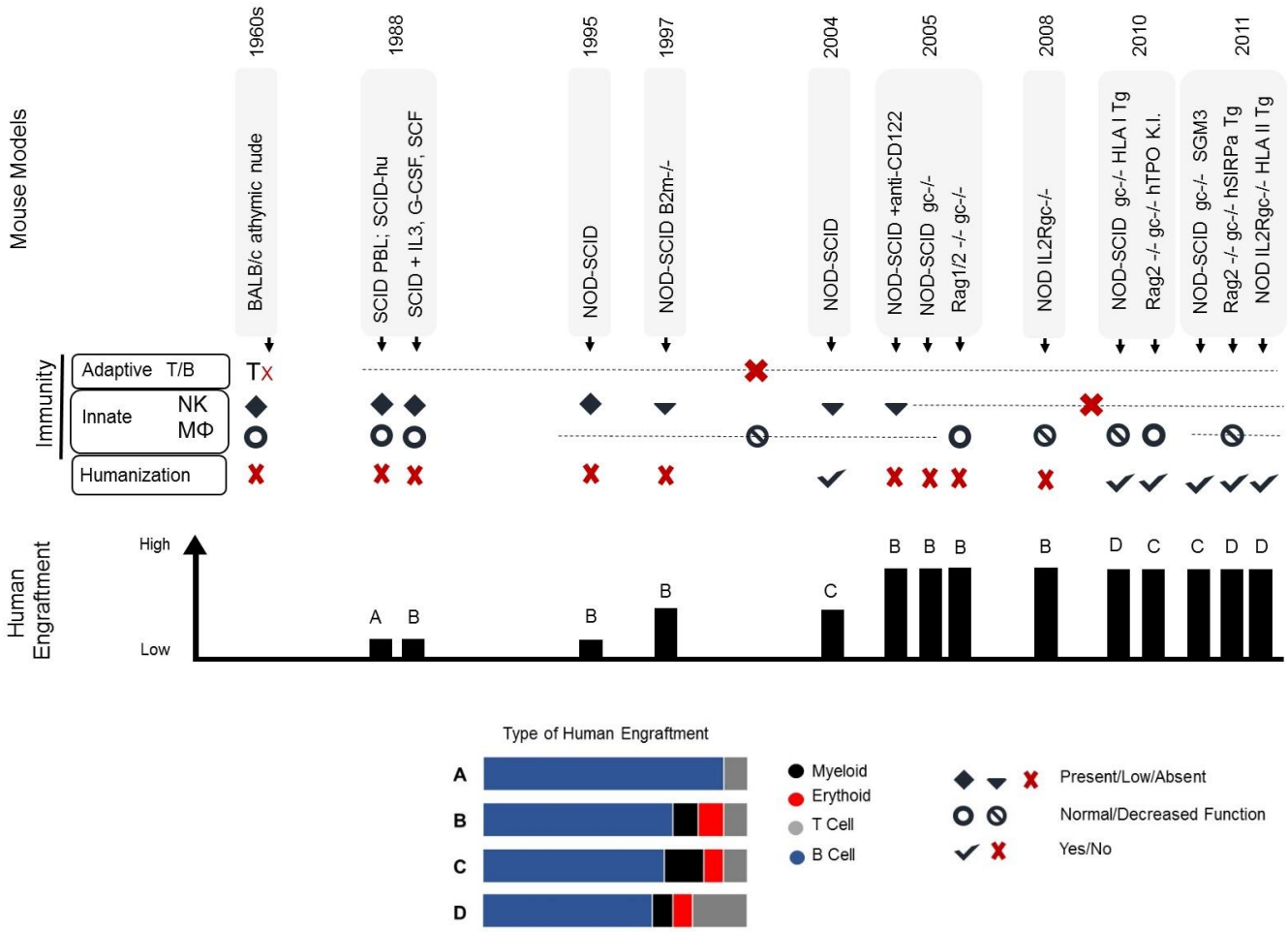


Figure 44: Timeline for the Development of Immune-Deficient Mouse Models.
Top half of the schematics shows the various mouse models developed since 1960s. Middle half of the diagram shows affected immune system in the mouse models with or without humanized element. Bar chart (bottom half) represents the level of human cell engraftment observed in the various mouse models. Alphabets present on top of each bar represents the type of engraftment (details underneath the bar chart).

The development of xenogeneic models (Figure 44) has enabled the rapid and facile *in-vivo* assessment of the patient's tumour tissues in immunocompromised mice. The first xenogeneic model developed in the late 1960s was the T-cell deficient BALB/c athymic nude (nude) mouse. In this model system, a single homozygous gene mutation in *Foxn*1 gene or the HNF-3/forkhead homolog 11 gene leads to the development of a hairless athymic (lack of functional thymus) mouse (Nehls et al., 1994). The immune system of nude mice is mainly characterized by a small population

of mature T cells, an antibody response confined to IgM class and an increased NK cell response. In these animals, absence of T cells allowed engraftment of some human tissues and tumour cells; however, the presence of mouse B cells and NK cells prevents human hematopoietic cell engraftment. The establishment of the severe combined immunodeficiency (SCID) mouse on a CB17 background (Bosma and Carroll, 1991), was a major advance which not only allowed the efficient transplantation and propagation of human tumour tissues but also enabled engraftment of the human haematopoietic cells in the host. These mice lack adaptive immunity due to a homozygous deletion in an essential enzyme namely Prkdc (Bosma et al., 1983). This DNA-dependent protein kinase enzyme is essential for repairing non-homologous ends of double-stranded DNA breaks following VDJ recombination in T- and B-cells. The defect in this enzyme impairs the VDJ recombination in these mice and therefore mice do not respond to either T- or B-cell mitogens. The humoral and cellular immune systems fail to mature in SCID mice; however, the spontaneous generation of mouse T- and B-cells during aging (also known as leakiness) and high levels of host NK cell and other innate immune activity remains the limiting factors. As a result, SCID mice often require conditioning with sublethal irradiation to engraft primary patient derived leukaemia cells.

The SCID-beige mouse model was developed by introducing a mutation in the lysosome trafficking regulating (Lyst) protein in the *Prkdc*^{scid} mutant mice, on a Balb/corCB57/BL/6J background (Mosier et al., 1993). As a result, these mice not only lacked T and B cells but also developed neutropenia as well as had reduced NK activity. Moreover, these mice also avoided the late immune-reconstitution observed in SCID mice. RAG mice strain was developed by targeting either rag1 or rag2 recombination activating genes. The RAG knockout mouse model harbours a germline mutation in which a large portion of the *rag* coding region is deleted (Shultz et al., 2000). These genes are DNA recombinases that are involved in recognizing and cleaving signal specific sequences for somatic rearrangements of B and T cell receptors. RAG mice strain do not show the leakiness or radio sensitivity as was observed in the SCID mice; however, these mice retained high levels of NK cell activity and show limited engraftment of human bone marrow stem cells.

Further attempts were made to manipulate the SCID mouse model to increase the human haematopoietic cell engraftment which led to the development of the non-obese diabetic-SCID mice. This hybrid model was developed by crossing the SCID mice (*Prkdc*^{scid}) to the non-obese diabetic (NOD) mouse background, resulting in immunological multidysfunction, including impaired NK cell function, reductions in macrophage function and complement-dependent hemolytic activity (Shultz et al., 1995). NOD mice harbour an essential as well as unique major histocompatibility complex (MHC) haplotype, namely H-2^{g}; as well as a polymorphism in the CTLA-4 and TNFα gene. All these factors together contribute to dysfunctional innate immune system observed in this mouse model. NOD-SCID mice showed marked improvement in the engraftment of the human haematopoietic cells in general, including primary bone marrow cells. Moreover, NK cell activity reduced significantly in this hybrid model and was lower in NOD-SCID mice than in CB17-SCID mice (Ito et al., 2002; Lowry et al., 1996). Lack of NK cell activity also reduced the requirement for conditioning with radiation prior to leukemic cell administration, however, conditioning with irradiation is still used by investigators. These mice have a propensity to developing thymic lymphomas. This mouse model enhanced the engraftability of the primary human cells including PBMC (Hesselton et al., 1995) as well as HSC (Lowry et al., 1996); and also enabled the maintenance of the leukemic characteristics of the tumour cells such as abnormal karyotype, stable antigenic markers and T- or B-cell receptor rearrangements. However, the use of this hybrid NOD-SCID mice model is limited due to relatively short life span (median survival of 257 days) and the residual activity of the NK cells as well as the other components of the innate immune system, thereby impeding the engraftment of the human lymphoid compartment.

The development of NOD/SCID/IL2rg$^{-/-}$ (or NSG) mouse which was developed by crossing the common-gamma-chain (interleukin 2 receptor gamma, IL2rg$^{-/-}$) knockout to the NOD-SCID mouse (Ishikawa et al., 2005; Ito et al., 2002; Shultz et al., 2005). IL2rg$^{-/-}$ mice produced independently in the early 1995 harbour homozygous mutation at the interleukin-2 receptor (IL-2R) γ-locus IL2rg (also known as common cytokine-receptor γ- chain) that is a crucial signalling component of the high-affinity receptors for IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21 (Ito et al., 2002; Traggiai et al., 2004). The lack of the IL-2R γ-chain in these hybrid mice leads to the multiple defects in both the

innate as well as the adaptive immune system, and prevents NK cell development. This hybrid model is resistant to developing lymphomas, therefore allowing for long-term human cell engraftment experiments (median survival >89 weeks). Depending upon the type of the host background, these hybrid mice strains of *Il2rg*$^{-/-}$ mice also include NOD.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Wjl} mice (NOD/LtSz-*scid Il2rg*$^{-/-}$), NODShi.Cg-*Prkdc*^{scid}*Il2rg*^{tm1Sug} mice (NOD/Shi-*scid Il2rg*$^{-/-}$), C.Cg-*Rag2*^{tm1Fwa}*Il2rg*^{tm1Sug} mice (BALB/c-*Rag2*$^{-/-}$*Il2rg*$^{-/-}$) and (H2d)-*Rag2*^{tm1Fwa}*Il2rg*^{tm1Krf} mice (H^{2d}-*Rag2*$^{-/-}$*Il2rg*$^{-/-}$) (Doulatov et al., 2010; Shultz et al., 2007). A common feature of all of the hybrid *Il2rg*–/– mice strains is that the host must be conditioned with sublethal γ-radiation before cell injection for optimal engraftment of the human cells.

Engraftment of the human tissues including human PBMCs and stem/progenitor cells in these hybrid strains of mice harbouring the IL2rg- mutations is greater and more reproducible than in all the other mice models described previously. For the first time, transplantation of human haematopoietic stem cells in NODSCID-IL2rg$^{-/-}$ mice showed engraftment of well differentiated progenitor cell populations into many lineages of the haematopoietic cells, including platelets, red blood cells and lymphoid compartment (both T- and B- cell) that progress through the expected stages of intrathymic development (Ishikawa et al., 2005; Shultz et al., 2005). As human cancer cells are transplanted into increasingly immunocompromised NSG mice, fewer and fewer cells are required to transfer the disease in such hybrid models, therefore making them ideal to use for studying the clonal evolution (Bonnet and Dick, 1997; Feuring-Buske et al., 2003; Quintana et al., 2008).

The landscape of somatic mutations in MDS patients with ring sideroblasts (MDS-RS) (Mufti et al., 2008) has become increasingly well defined (Makishima et al., 2012; Mian et al., 2013; Papaemmanuil et al., 2011; Visconte et al., 2012b; Yoshida et al., 2011b). To date, functional and mutational evaluations of hematopoietic cells in MDS in general, have predominantly relied on total nucleated cells or purified CD34^{+} bone marrow cells, which have limited power to identify cell lineage specific genomic aberrations. Specifically, the origin of *SF3B1* mutations, the clonal composition, evolution as well as the engraftment kinetics of the haematopoietic progenitors that carry the *SF3B1* mutations remain unknown. In addition, the specific step within the developmental schema at which a clone attains a particular genetic aberration

necessary to emerge or re-emerge as a dominant clone remains unknown. For instance, the sequential acquisition of oncogenic genetic alterations (such as *FLT3* and *RUNX1*) in *SF3B1* mutant MDS-RS patients results in disease progression to acute myeloid leukaemia (Mian et al., 2013). However, the order in which these 'oncogenic' mutations are acquired and the cell type in which these mutations arise in MDS-RS have yet to be determined.

Over the years, it has been reported that self-renewing HSCs continuously acquire somatic aberrations, while most of them are passenger mutations, some 'potent mutations' can constitute a reservoir of pre-leukemic stem cells (Greaves and Maley, 2012; Shackleton et al., 2009; Swanton, 2012). The first study to report clonal spectrum at single-cell level through multiplex FISH analysis was in childhood ALL (Anderson et al., 2011). However, the recent developments of genomic technologies, stem cell isolation as well as xenotransplantation models has started to lead to a better understanding of the complex clonal architecture and mutational hierarchy of phenotypically and functionally defined 'malignant stem cells' in AML (Klco et al., 2014).

Therefore, we hypothesized that NSG IL2rg$^{-/-}$ xenograft mouse model and single cell analysis would be attractive methods to investigate the *SF3B1* mutant bone marrow cells. This in combination with FACS analysis, whole-exome sequencing and targeted mutational analysis were used in this study to determine the origins of *SF3B1* mutations as well as depict the clonal spectrum and evolution seen in the bone marrow compartment of the MDS-RS patients.

5.2 Results

5.2.1 *SF3B1* mutation originates in a cell with an HSC phenotype and persists in mature myeloid progenitors

WES data from bone marrow $CD34^+$ cells from a cohort of 11 MDS-RS and 1 congenital sideroblastic anaemia patient was available as described previously in section 3.1. All the patients had *SF3B1* mutations (Lys700Glu, n=4; His662Gln, n=3; Glu622Asp, n=2 and Asp781Gly, n=1; Table 32). All the *SF3B1* mutations were heterozygous in nature. Bone marrow cells from up to 6 patients were used for stem cell isolation as well as a number of *in-vitro* and *in-vivo* analyses (Table 32 and Table 34). Three additional patients were used for single cell clonogenic assays (CFC) assays. A constitutional *ALAS2* (Arg425Cys) gene mutation (Camaschella, 2009; Cotter et al., 1999; Kaneko et al., 2013) was detected in the patient with congenital sideroblastic anaemia, but no other somatic mutations including *SF3B1* were observed in this case. Thus, this patient served as an ideal control for all the subsequent *in-vivo* assays. We also used 3 haematologically normal controls for in-vivo assays (Table 34).

Table 32: Clinical characteristics of patients analysed.

Sequential bone marrow samples were available from 2 patients as described in the table. TP-Time point, M-Male, F-Female, A-Alive, D-Dead. †Immunophenotyping demonstrates 2 separate populations of $CD34^+$ cells: (1) CD34+ (bright) ; CD13+ ; CD33+ ; weak CD117+ ; ~1/2 MPO+ ; TdT+ ; CD14- ; CD19- ; CD7- ; CD79a- ; cyt.CD3- (~20% TNCs). (2) CD34+ ; CD13+ ; CD33+ ; CD117- ; ~1/2 MPO+ ; TdT- ; CD14+ ; CD19- ; CD7- ; CD79a- ; cyt.CD3- (~60% TNCs).

Patient ID	Sample Time Point	Age	Sex	WHO Diagnosis	Treatment Prior To Sample	BM Blasts (%)	Hb	NPL	Platelet Counts	Ringed Sideroblast %	IPSS	Cytogenetics	Transfusion Dependant	Survival Status	AML Evolution	Disease Progression
MDS 1	TP 1 (Day 0)	77	M	RARS-T	Epo/ GCSF	2	8.8	9.53	268	>15	Low	NORMAL	Yes	A	No	No
MDS 1	TP2 (Month 18)	-		RARS-T	Epo/ Len	2	10.5	1.9	248	>15	Low	NORMAL	Yes	A	No	No
MDS 2	TP1 (Day0)	75	M	RCMD-RS	Epo/ GCSF	1	8.6	1.87	166	41	Low	NORMAL	Yes	A	No	Yes
MDS 2	TP2 (Month 18)	-	-	RCMD-RS	Epo/ Len	1	7.1	3.67	137	50	-	NORMAL	Yes	A	No	Yes
MDS 2	TP3 Month 29)	-	-	sAML	Epo/ Len	90 †	7.5	1.63	29	-	-	46,XY,add(7)(q32),add(10)(p13)[7]/46,XY, add(7)(q22),add(7)(q32),add(10)(p13)[3]	Yes	D	Yes	Yes
MDS 3	N/A	63	M	RARS	Epo/ GCSF	1	9.1	3.15	570	>15	Low	NORMAL	No	A	No	No
MDS 4	N/A	58	F	RARS	None	1	10.7	3.21	196	24	Low	NORMAL	No	A	No	No
MDS 6	N/A	74	M	RCMD-RS	Epo/ Len/ Aza	1	12.1	3.44	151	35	Low	NORMAL	No	A	No	No
MDS 5	N/A	22	M	Congenital sideroblastic anaemia	None	1	10.6	4.7	400	32	N/A	NORMAL	Yes	A	No	No

In order to identify the target cell within the hematopoietic compartment, which carries the *SF3B1* mutation HSCs, MPPs, GMPs and MEPs were isolated from the bone marrow of 4 patients (MDS1, MDS2, MDS4 and MDS6; Figure 45, Appendix Figure 1 and Table 34).

The cell surface antigens present in our antibody panel and used for separation of specific stem as well as progenitor cell compartments (Table 16) were similar to the ones used in previous studies (Pang et al., 2013). In addition, CD49f cell surface antigen was used to differentiate between HSCs and MPPs. CD49f (integrin 6) expressed within $Lin^{-}CD34^{+}CD38^{-}CD45RA^{-}$ population have been shown to efficiently generate long-term multi-lineage grafts in *in-vivo and* the loss of CD49f expression identified transiently engrafting MPPs (Notta et al., 2011). Initially the antibody panel was optimised using bone marrow cells obtained from 'haematologically normal' subjects as depicted in Figure 45. Subsequently, bone marrow cells from MDS-RS patients were prepared and sorted on a BD FACS Aria SORP (San Jose, CA) operating in 4-way purity sort mode, and collected into 1.5 ml microfuge tubes (see methods chapter for full details). After sorting, the purity of MPPs, GMPs and MEPs were checked and a sorting purity of ≥97% was routinely obtained. Cell number obtained for each compartment for all the patients are mentioned in Figure 46.

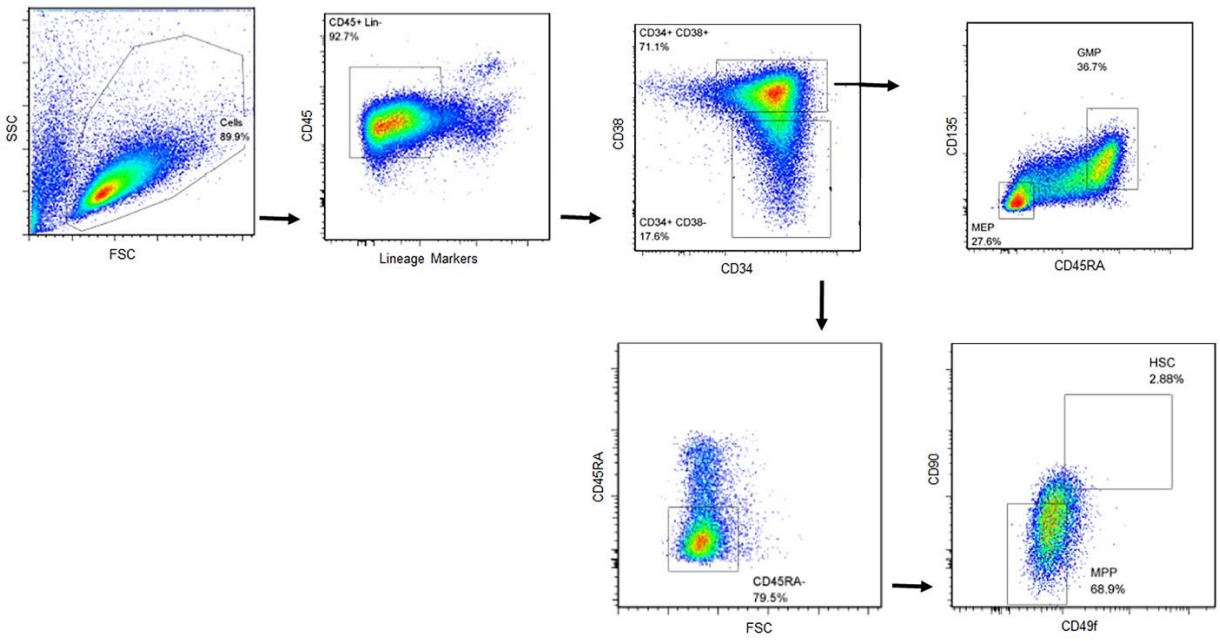


Figure 45: Representation of the isolated hematopoietic stem cell subpopulations.
Viable lineage negative bone marrow CD34$^+$ cells from 'haematologically normal' control were defined by hierarchical gating on a combination of forward (FSC) and side scatter (SSC), propidium iodide (not shown), CD45 and a lineage cocktail. These CD34$^+$ cells were further gated first into CD34$^+$CD38$^-$ and CD34$^+$CD38$^+$ populations and then subdivided into either HSCs (CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$CD49f$^+$) and MPPs (CD34$^+$CD38$^-$CD90$^-$CD45RA$^-$CD49f$^-$) or into GMPs (CD34$^+$CD38$^+$CD135$^+$CD45RA$^+$) and MEPs (CD34$^+$CD38$^+$CD135$^-$CD45RA).

Following the cell separation, targeted mutational analysis specific to patients' *SF3B1* mutations was performed on haematopoietic cell fractions (HSCs, MPPs, GMPs and MEPs). Using *SF3B1* mutation specific primers, PCR was directly performed on 30-50% of the FACS purified cells from each compartment with subsequent sequencing on MiSeq platform and/or Sanger sequencing (see methods section 2.2.6 for more details). Sequencing data revealed that patient specific *SF3B1* mutation were present in HSC population (Figure 46). *SF3B1* mutations were also detected in MPPs, GMP and MEP fractions from all patients (Figure 46). All the *SF3B1* mutation detected in patients progenitor compartment were heterozygous in nature. Notably, the MAB was largely maintained between the HSCs, MPPs and more differentiated myeloid progenitors and was similar to bone marrow CD34$^+$ cells (pooled) as shown in Figure 46.

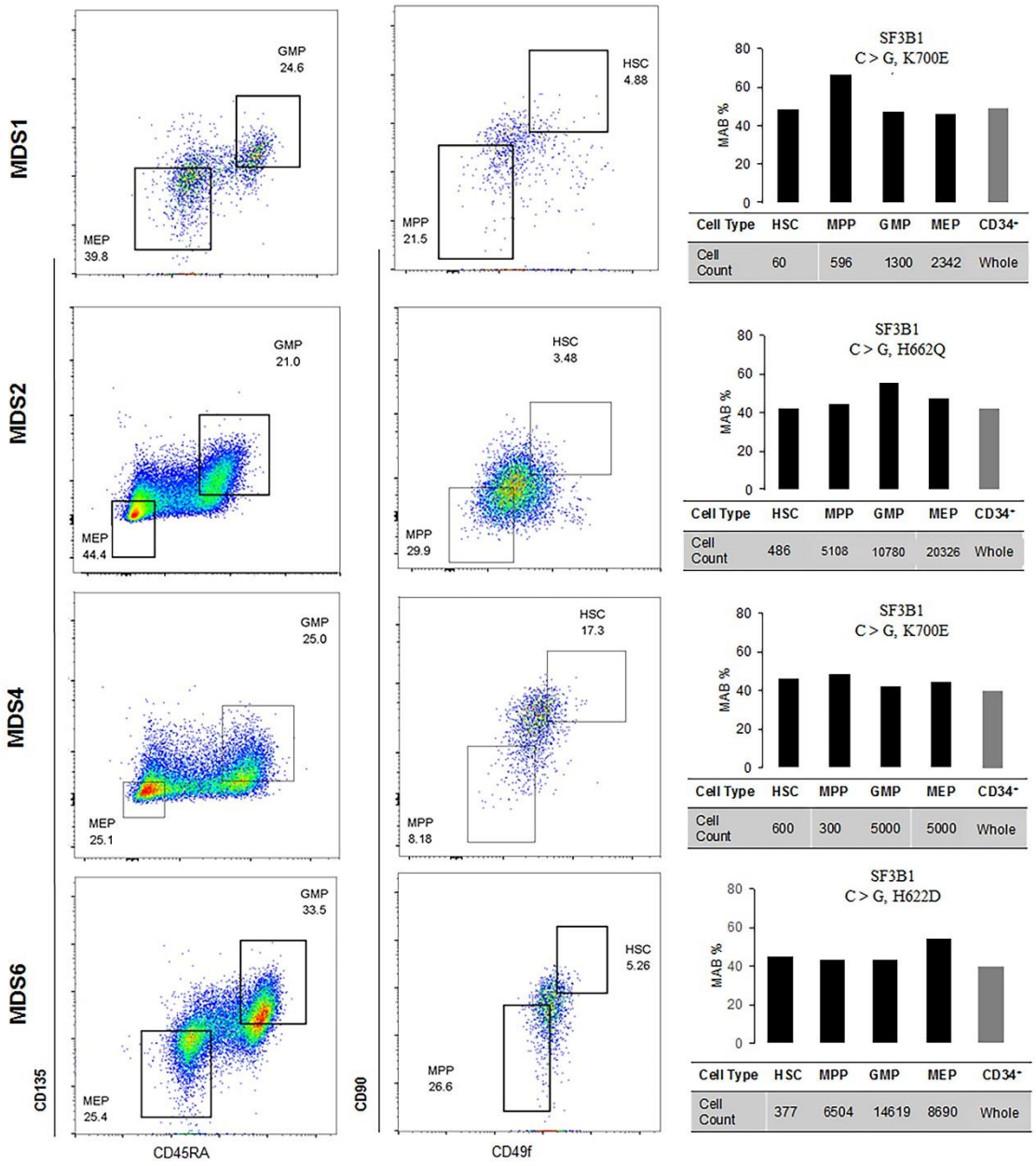


Figure 46: *SF3B1* mutations propagates from HSCs to their progeny.

Representation of the isolated haematopoietic stem cell subpopulations. Numbers present within the FACS plots represent the percentage of the cells of each progenitor compartment. *SF3B1* mutations occurs in rare HSC phenotypic cells and largely maintained in more committed progenitors MPPs as well as in more differentiated progenitors GMPs and MEPs. Tables show the progenitor cell counts for each patient sample. Bar charts indicate the *SF3B1* MAB for respective cell fraction obtained through MiSeq sequencing platform. The cell count is based on the number of events detected by the FACS during cell sorting. The average sequencing coverage across all amplicons was ≥800 reads. MAB for each cell population was confirmed by independent PCRs. MAB- mutant allele burden.

5.2.2 *SF3B1* mutations persist throughout myeloid but not lymphoid differentiation

In order to determine the *SF3B1* mutation status in lymphoid cell and mature erythroid cell compartment of the MDS-RS patients, $CD34^-CD3^+$ T-cells, $CD34^-CD3^+CD4^+$ T-cells , $CD34^-CD19^+$ B-cells, $CD34^-CD71^+CD235^+$ erythroblasts and $CD34^-CD3^-CD235^-$ were isolated from total bone marrow cells using BD FACS Aria SORP (San Jose, CA; see methods section 2.2.8 for more details). Initially, three patients in whom *SF3B1* mutations were detected in total bone marrow cells were used for this analysis. Using the appropriate flow cytometry values, the cells of interest were then gated by the following cell surface markers as described in methods chapter. After sorting, the purity of all the cell fractions was checked and a sorting purity of ≥97% was routinely obtained.

Following the cell separation, targeted mutational analysis specific to patients' *SF3B1* mutations was performed on DNA isolated from all cell fractions. PCR was performed using *SF3B1* mutation specific primers and subsequently sequenced using Sanger sequencing (see methods section 2.2.7 for more details). Mutation load was also checked in total $CD34^+$, $CD34^-CD71^+CD235^+$ erythroblasts and $CD34^-CD3^-CD235^-$ cells showed approximately the same *SF3B1* MAB (according to Sanger sequencing) as seen for paired total bone marrow and $CD34^+$ cells (Figure 47). Importantly, no *SF3B1* mutation was detected above background detection levels in $CD3^+$ T-cells, $CD3^+CD4^+$ T-cells or $CD19^+$ B-cells according to Sanger sequencing (3/3 mutant patients), indicating that the mutant clone has a growth advantage only in the myeloid lineage in MDS-RS (Figure 47).

In addition, high-depth sequencing using MiSeq platform (Second-generation sequencing) performed on $CD3^+$ T-cells from one additional patient (MDS1) confirmed the absence (MAB 0%, sequencing read depth 7689X) of *SF3B1* mutation in T-cells.

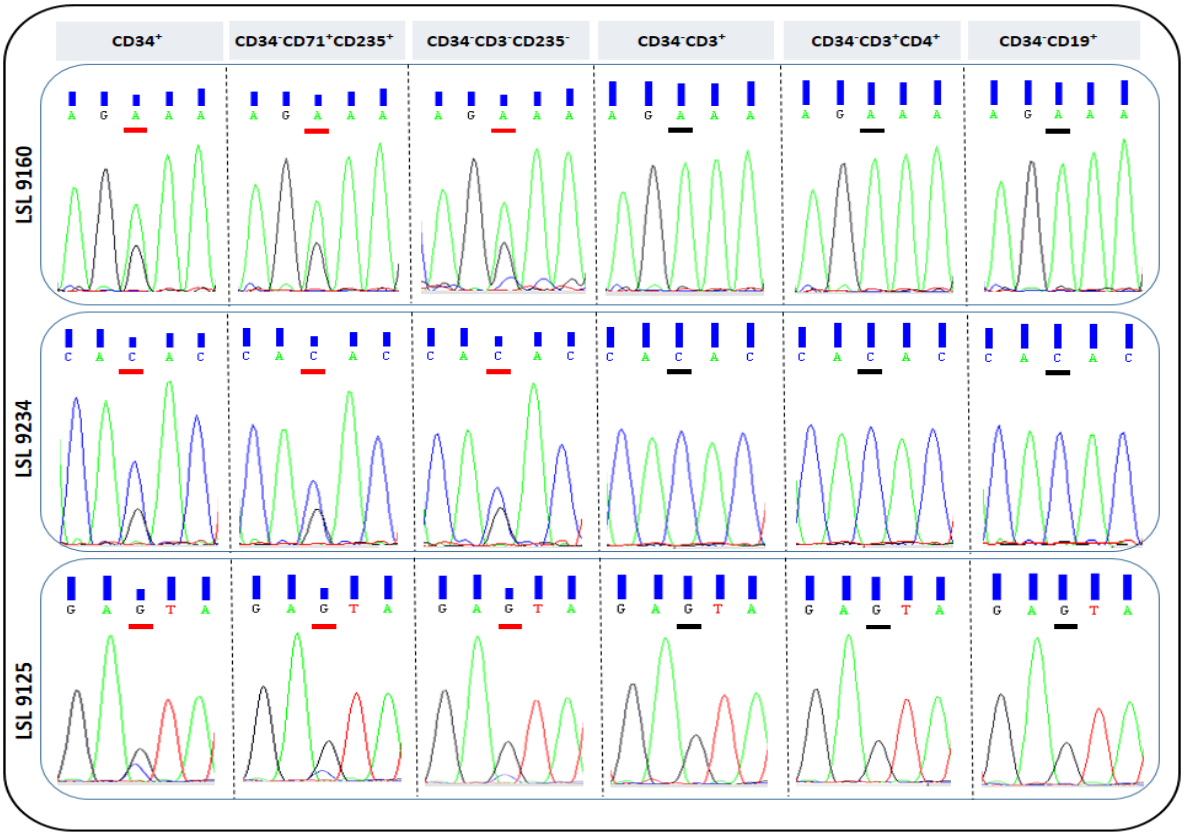


Figure 47: *SF3B1* mutation persists in the myeloid but not in lymphoid lineage.

Sequencing analysis of myeloid and lymphoid cells from 3 RARS patients with *SF3B1* mutations. The three horizontal rows represent the 3 patient samples and the vertical columns represent the cell types isolated from the bone marrow. $CD34^{-}CD71^{+}CD235^{+}$ erythroblasts and $CD34^{-}CD3^{-}CD235^{-}$ cells showed approximately the same *SF3B1* mutant allele burden as seen in the $CD34^{+}$ cells. No *SF3B1* mutation was detected above background in $CD3^{+}$ T-cells, $CD3^{+}CD4^{+}$ T-cells or $CD19^{+}$ B-cells.

5.2.3 WGA maintains the allelic burdens across the genome

Technical difficulties in isolating and analysing rare haematopoietic stem cells and/or small number of cells (<1000 cells) have made limited progress to date. However, the recent developments of genomic technologies, stem cell isolation as well as high depth mutational analysis has started to lead to a better understanding of the complex clonal architecture (Klco et al., 2014), especially using small cell numbers. In this study, we had to perform genomic analysis on isolated stem/ progenitor cells and small numbers of cells (obtained from clonogenic assays and mice experiments). Therefore, we used whole genome amplification technology to increase the amount of DNA which could be used for downstream sequencing analysis. WGA of cells was performed with GenomePlex Single Cell Whole genome amplification kit (WGA4, Sigma-Aldrich Co, LLC) and/or GenomePlex SeqPlex XE kit (SEQXE, Sigma-Aldrich Co, LLC). WGA amplification was performed according to the instructions of the manufacturer (as detailed in methods chapter), along with a no cell reaction as a negative control and a reaction of human tissue genomic DNA as positive control.

The degree of quantitative accuracy of the allele burdens for WGA DNA was validated independently by performing whole-exome sequencing on paired gDNA (non-WGA) and WGA DNA from the samples (Figure 48). LOH was mapped from WES data according to the following criteria: Variants found in 1000 genome project database with a population allele frequency less than 0.001 were selected and filtered out for known genomic duplicated regions, read depth <20 and somatic status according to VarScan (see methods section 2.2.4.6 and 2.2.4.8 for more details). The filtered set of variants were then plotted graphically for the required chromosome. Following the primary analysis, similar distribution of known SNPs across the exome between the gDNA (non-WGA) and WGA DNA was observed (Figure 48A). To validate this further, mutational screening for 10 genes on paired gDNA (non-WGA) and WGA DNA from 2 samples (MDS1 and MDS2) was performed on independent patients (Figure 48B). MAB for specific mutations was also shown to be comparable between gDNA (non-WGA) and WGA DNA in targeted amplicon sequencing in these control experiment.

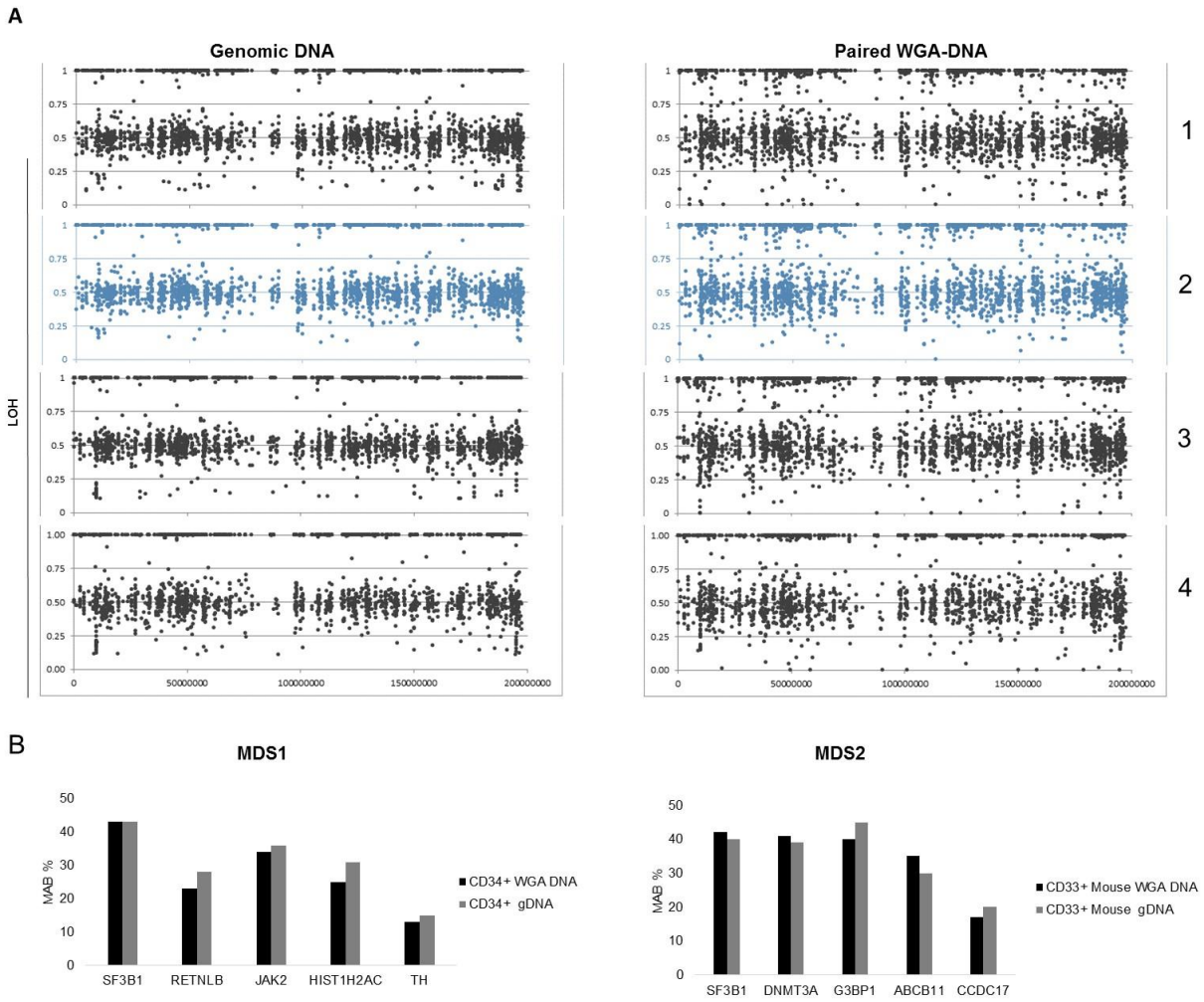


Figure 48: Validation of gDNA vs WGA-DNA through SNP profiling by WES and targeted-mutational analysis.

(A) Comparison of SNV profiles from control exome experiments based on gDNA vs WGA-DNA. Paired exome experiments (1-4) of gDNA (non-amplified) vs WGA-DNA were processed and analysed in VarScan. Variants were selected and plotted based on being called germline or LOH in VarScan, due to presence at similar levels in both samples of a pair or significant allele burden (AB) change between the two respectively, had >20 read depth and had >10% AB in unamplified experiment data. Chromosome 3 data is exampled here as representative Figure. Panel 1 and panel 2 are based on the experiment of human $CD33^+$ cells isolated post mouse engraftment. Panel 3 is based on primary patient $CD34^+$ cells. Panel 4 is based on patient skin material. Panel 1, panel 3 and panel 4 show all variations as detailed above. Panel 2 shows sub-fraction of panel 1 that is found in dbSNP137. (B) Comparison of somatic mutations between paired gDNA vs WGA-DNA. Mutant allele burden of 5 gene mutations in primary $CD34^+$ bone marrow and $CD33^+$ HEC cells from MDS1 and MDS 2. MAB- Mutant Allele Burden, gDNA- Genomic DNA, WGA-Whole-genome amplified DNA.

5.2.4 SF3B1 mutant cells maintain their differentiation and proliferation potential *in-vitro*

To evaluate the *ex-vivo* clonal growth potential of *SF3B1* mutant bone marrow $CD34^+$ cells from 6 MDS-RS patients (MDS1, MDS2, MDS3, MDS4, MDS6 and MDS142), one congenital sideroblastic anaemia patient (MDS6) and 3 'haematologically normal' controls were plated in methylcellulose to assess their colony-forming cell abilities. Cells were cultured for 14 days. In addition, 3 patients (MDS1, MDS2, MDS3) were also grown in liquid culture (long-term culture, or LTC; See methods chapter for more details) to assess their long term proliferation potential.

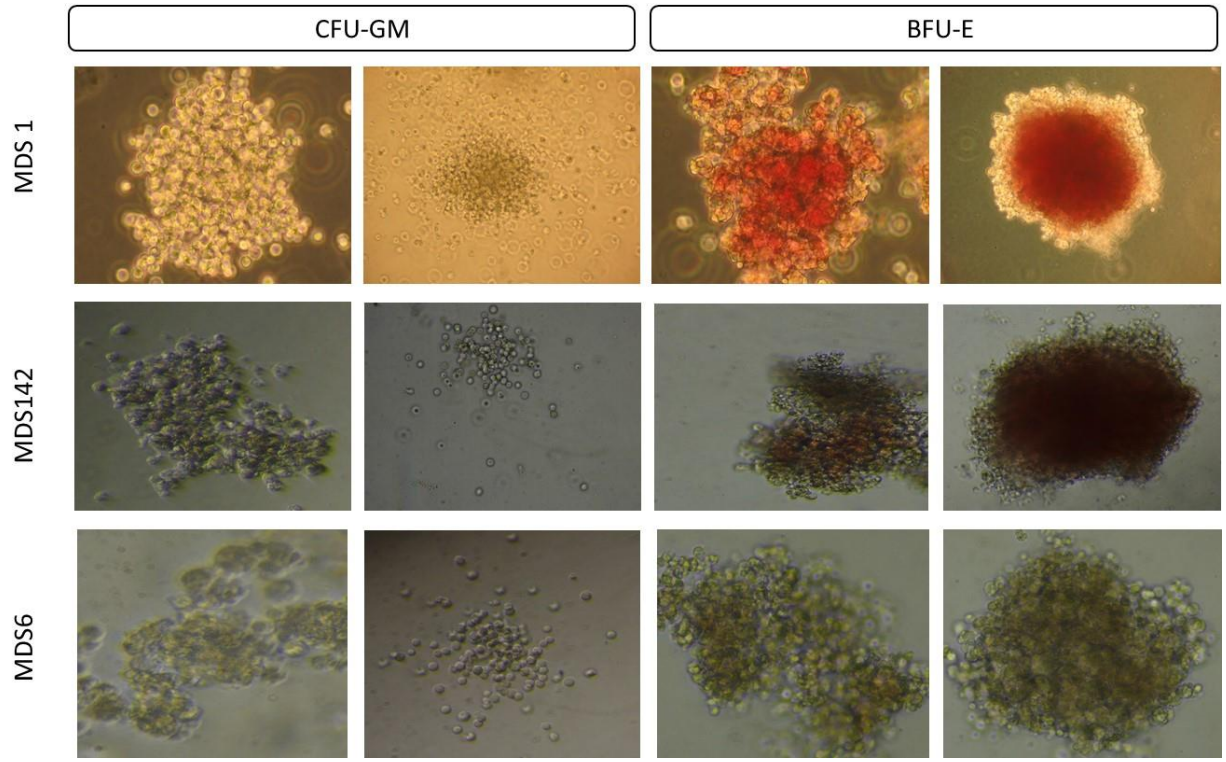


Figure 49: CFC-GM and BFU-E colonies.

Cells cultured from CD34+ bone marrow cells of MDS-RS patients after 14 days of culture in methylcellulose-based medium. Each colony was picked and processed for mutational analysis. Cultured colonies were analysed based only on the granulocyte/macrophage and erythroid selection criteria.

Overall no preferential bias towards any particular cell lineages in the short-term CFC culture was observed (as exampled in Figure 49 for three patients; Table 33) as compared to haematologically normal controls. MDS-RS patients with *SF3B1* mutations generated equal number of CFC-GM and BFU-Es. However, MDS4 yielded low colony numbers as compared to other patients. Patient with congenital sideroblastic anaemia had more CFC-GM compared to BFU-Es.

After 14 days culture, colonies were picked, harvested and washed twice with PBS. DNA was isolated from the pooled colonies from three patients (MDS1, MDS2 and MDS3). In addition individual colonies collected were used in single cell colonogenic mutational analysis (discussed in detail section 5.2.9).

Table 33: Colony-forming cell (CFC) assay analysis.

Analysis of the differentiation capacity of CD34+ cells from 7 MDS-RS patients, one congential sideroblastic anaemia and 3 haematologically normal controls. The table show the numbers of granulocyte-macrophage (CFU-GM) and erythroid (BFU-E) colonies generated by total $CD34^+$ cells in the presence of cytokines IL-3, Tpo and Epo, respectively. Cells were plated in triplicate. Data are mean ± SD. CFU-GM and BFU-E frequency for some of the patients are shown as total colony numbers. † Congential sideroblastic anaemia

Experiment no	Number of $CD34^+$ cells seeded	Number of CFU-GM colonies at day 14	Number of BFU-E colonies at day 14
MDS1	500	27.7 ±6.3	23.7 ±0.9
MDS2	500	71.7 ±11.4	
MDS3	500	37 ±10.6	34.3 ±3.9
MDS4	500	5.3 ±3.3	7.3 ±0.9
MDS5†	500	54.0 ±6.2	27 ±1.6
Normal BM1	500	31 ±7.9	35.3 ± 3.1
Normal BM2	500	35.7 ± 7.5	29.3 ±11.3
Normal BM3	500	37 ±6.5	37 ±7.5

SF3B1 mutated cells also maintained their proliferation potential in the LTC conditions. After 5 weeks culture, live cells were collected, stained using human $CD45^+$ cell surface antigen and sorted by using the BD FACS Aria SORP (See methods section 2.2.8 for more details). Sorted cells were washed in PBS, harvested and stored as cell pellet. DNA was also isolated from LTC cultured cells (MDS1, MDS2 and MDS3).

Following on, targeted mutational analysis specific to patients' *SF3B1* mutations was performed on the DNA isolated from pooled-CFC and LTC derived cells (MDS1, MDS2 and MDS3). Using *SF3B1* mutation specific primers, all the samples were PCR amplified and subsequently sequenced using MiSeq/Roche 454-sequencing sequencing (See methods section 2.2.5 and 2.2.6 for more details). Sequencing data showed that *SF3B1* mutations observed in patients $CD34^+$ cells were present and maintained in both CFC and LTC from all 3 patients (as exampled for one patient MDS1 in Figure 50, data for MDS2 and MDS3 are shown in Figure 54). All the mutation were heterozygous aberrations (MAB up to 50%). These findings suggest that, in this subgroup of patients, the $CD34^+$ *SF3B1* mutant cells are clonal neoplastic cells with the ability to differentiate into mature erythroid and granulocytic colonies as well as maintain their long term proliferating potential.

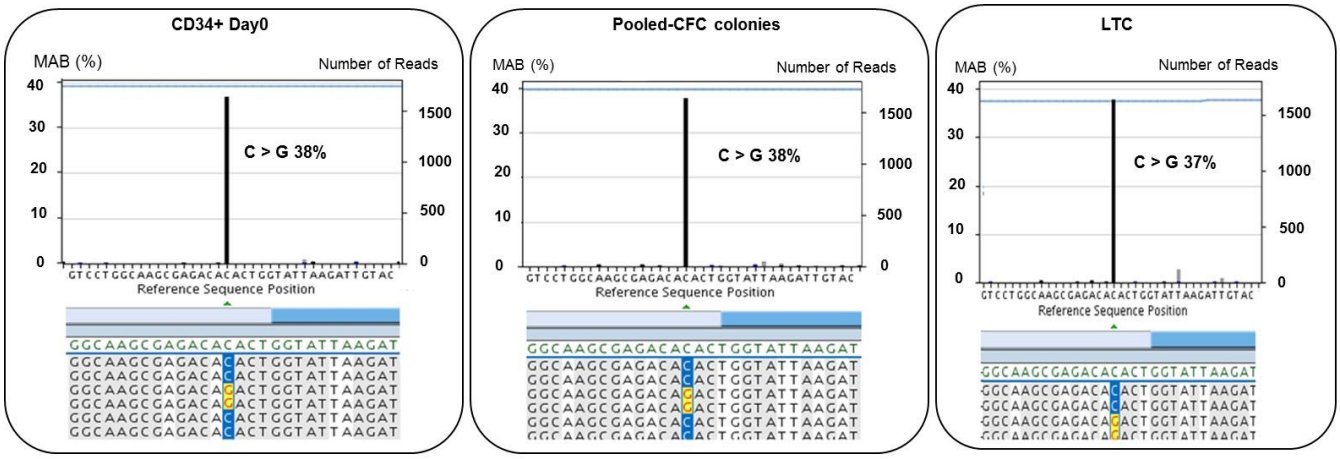


Figure 50: *SF3B1* mutations are maintained in CFC and LTC culture conditions.
Bone marrows cell from MDS1 patient were subjected to CFC and LTC culture conditions. Mutational analysis on the pooled CFC and LTC cultured cells show that the mutant allele burden (MAB) of *SF3B1* mutation remained almost identical.

5.2.5 *SF3B1* mutations persists in NOD/SCID/IL2rγ^{-mu} Mice (NSG)

NSG mice models in which primary human leukemic cells are transplanted have previously been used to expand our understanding of the pathogenesis of human leukaemia (Anderson et al., 2011; Bonnet and Dick, 1997). Although MDS patient cells are difficult to engraftment in xenograft models, the recent use of NSG mice as well as intra-bone injection (with or not the co-injection of mesenchymal stroma cells) have been reported to yield higher engraftment (Medyouf et al., 2014; Pang et al., 2013; Woll et al., 2014). For this reason, a murine xenograft model of *SF3B1*/MDS-RS using NSG mice was used to study the engraftment kinetics of *SF3B1*-mutant CD34$^+$ cells (Figure 51, Figure 52 and Table 35).

Next, CD34$^+$ bone marrow cells from four *SF3B1* mutated MDS-RS patients (MDS1, MDS2, MDS3 and MDS4) and one congenital sideroblastic anaemia patient (MDS5 with *ALAS1* mutation) were prepared (see methods section 2.2.8 for more details). The congenital and acquired sideroblastic anaemia's share the same phenotypic characteristics mainly characterized by the presence of ring sideroblasts in the bone marrow. Therefore, to compare the engraftment kinetics of *SF3B1* mutant haematopoietic stem cells, bone marrow cells from congenital sideroblastic anaemia patient served as an ideal control for *in-vivo* experiments. In order to assess whether the NSG mice biases the engraftment of any particular hematopoietic lineages (myeloid or lymphoid), we also injected 'haematologically normal' adult bone marrow CD34$^+$ cells from 3 healthy donors into NSG mice. Prior to transplantation, mice received a sub-lethal dose of radiation (330-375 cGy) from a cesium-137 source. Depending upon the total number of CD34$^+$ cells available from each patient, 0.65x10^5 to 2x10^5 bone marrow CD34$^+$ cells were injected into the tibia of NSG mice (Table 34).

Table 34: Patient characteristics and engraftment of bone marrow CD34$^+$ cells.
Bone marrow cells were obtained from MDS-RS patients and one patient with congenital sideroblastic anaemia. N-Normal, N/A- Not applicable.

Patient No	Diagnosis	Known Disease Related Mutations	Percentage of Ringed Sideroblast	Cytogenetics	WES	Stem Cell Analysis	LTC	CFC	IBM CD34+ Cells Injected in NSG Mice	No. of Mice Injected/Engrafted	Percentage of Human Cell Engraftment (Week 6)			Percentage of Human Cell Engraftment at Termination (week 18-20)		
											CD45+	CD45+ CD19+	CD45+ CD33+	CD45+	CD45+ CD19+	CD45+ CD33+
MDS1	RARS-T	*SF3B1* Lys700Glu	>15%	N	Yes	Yes	Yes	Yes	200,000	1/1	0.1	5.6	76.4	4.9	0	90
									200,000	1/1	1.1	19	75	3.6	0	96
									200,000	1/1	0.2	3.5	88	3.8	0	92
MDS2	RCMD-RS	*SF3B1* His662Gln	41%	N	Yes	Yes	Yes	Yes	67,000	1/3	0	0	0	0.1	5	95
MDS3	RARS	*SF3B1* His662Gln	>15%	N	Yes	No	Yes	Yes	100,000	1/1	1.7	64	19	1.1	0.2	97
									100,000	1/1	0.1	33	28	0.3	0	97
									100,000	1/1	1.4	53.2	26	2.4	0.2	95
MDS4	RARS	*SF3B1* Lys700Glu	24%	N	Yes	Yes	Yes	No	65,000	1/2	0	0	0	0.1	0	83
MDS6	RCMD-RS	*SF3B1* Glu622Asp	35%	N	Yes	Yes	No	No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MDS5	Congenital Sideroblastic anaemia	*ALAS2* Arg452Cys	32%	N	Yes	No	No	No	100,000	1/1	5.35	1.36	93	0.5	28.3	61
									100,000	1/1	1	2.43	94	0.3	3.17	89.3
Normal BM1	Haematologically Normal	N/A	N/A	N/A	No	No	No	No	100,000	1/1	N/A	N/A	N/A	2.87	87.00	12.2
									100,000	1/1	N/A	N/A	N/A	7.75	94.00	5.80
									100,000	1/1	N/A	N/A	N/A	9.9	89.00	10.00
Normal BM2	Haematologically Normal	N/A	N/A	N/A	No	No	No	No	100,000	1/1	N/A	N/A	N/A	1.4	94.00	4.30
									100,000	1/1	N/A	N/A	N/A	3.3	96.00	3.00
									100,000	1/1	N/A	N/A	N/A	3.6	92.00	6.60
									100,000	1/1	N/A	N/A	N/A	2.2	95.00	5.00
Normal BM3	Haematologically Normal	N/A	N/A	N/A	No	No	No	No	100,000	1/1	N/A	N/A	N/A	9.3	83.00	16.00
									100,000	1/1	N/A	N/A	N/A	21.2	56.00	43.00
									100,000	1/1	N/A	N/A	N/A	3.5	80.00	20.00
									100,000	1/1	N/A	N/A	N/A	9.8	84.00	15.00

Engraftment levels based on the recovery of human CD45$^+$ cells from engrafted mice at week 18-20 ranged between 0.1% and 4.9% (Figure 51A-51B and Table 34). Notably, mice injected with a high cell number from MDS-RS patients showed greater human cell engraftment. Further analysis of the bone marrow cells retrieved from engrafted mice using lineage-specific markers (hCD33 for myeloid cells and hCD19 for B-cells) revealed that the majority of the cells engrafted in mice were of myeloid lineage with little or no lymphoid cell engraftment. In contrast, mice injected with CD34$^+$ cells from a congenital sideroblastic anaemia patient with constitutional *ALAS* mutation but no *SF3B1* mutation showed myeloid as well as lymphoid cell engraftment (Figure 52A and Table 34). Similarly, engraftment observed in mice transplanted with CD34$^+$ cells from normal healthy controls exhibited a higher ratio of lymphoid/myeloid cell as compared to *SF3B1* mutant cases (Figure 52C). Therefore data from our study suggests that the presence of *SF3B1* mutations, perhaps along with additional 'cooperating factors', adversely affects the lymphoid cell development in this model, which is in agreement with the previous xenotransplantation models proposed elsewhere (Pang et al., 2013; Woll et al., 2014)

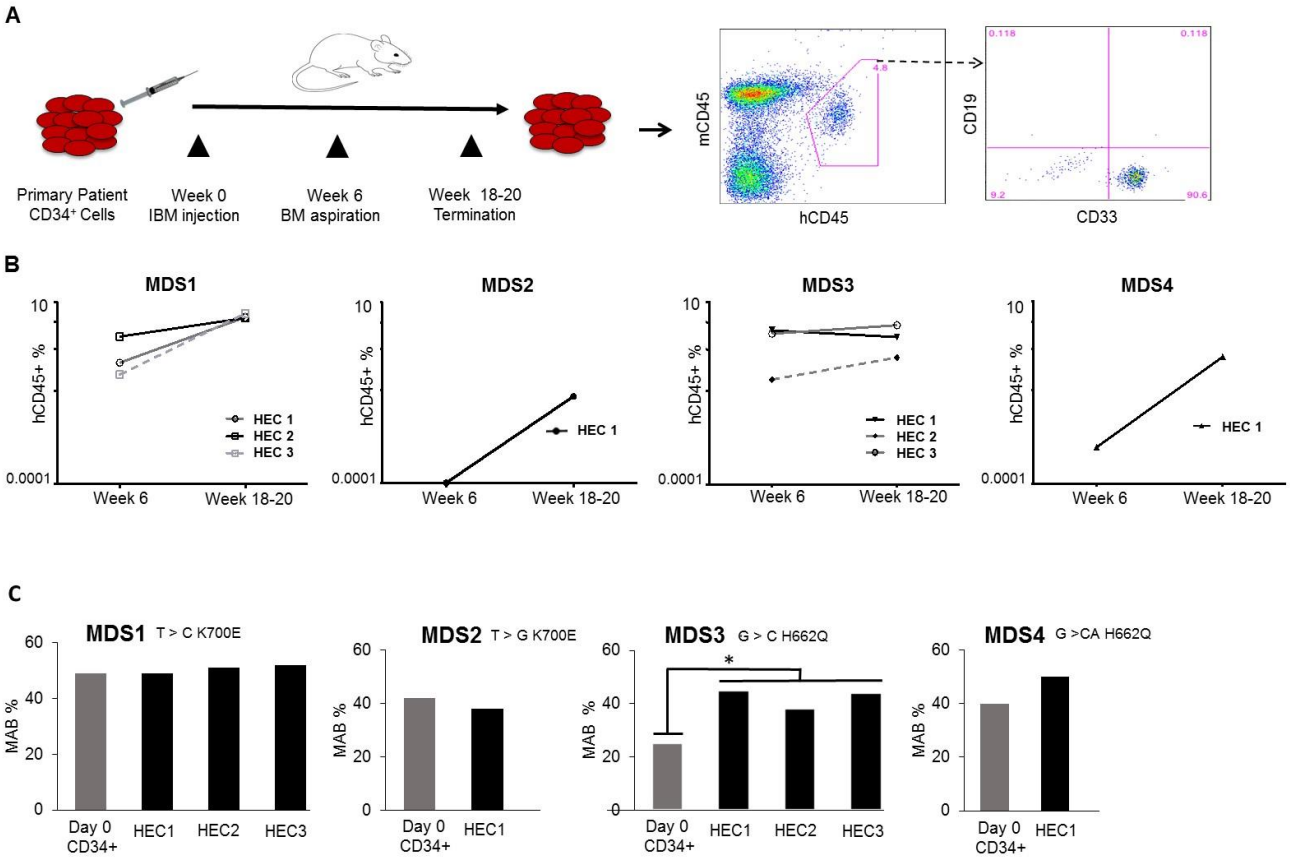


Figure 51: Engraftment and genotype of human RARS-originated and control hematopoietic cells in the bone marrow of NOD/SCID/IL2rγ−/− mice.

(A) Schematic representation of the xenograft model. Right hand side FACS panel shows the flow cytometric profile of bone marrow cells recovered from one human cell engrafted mice (HEC). The majority of human CD45 expressing cells were positive for a myeloid marker $CD33^+$ in all analysed cases in this study. (B) Percentage of human $CD45^+$ cells in the bone marrow of NSG mice at 6 and 18-20 weeks after transplantation (MDS1, n=3; MDS2, n=1; MDS3, n=3 and MDS4, n=1). (C) Targeted mutational analysis shows the presence of concordant SF3B1 mutations in primary $CD34^+$ bone marrow sample (grey) and xenograft (black) in all analysed cases. Three independent PCR/sequencing experiments were performed to confirm/determine the mutant allele burden throughout the experiments. The sequencing coverage across the *SF3B1* amplicons was ≥ 1000 reads. *t-test p-value <0.01

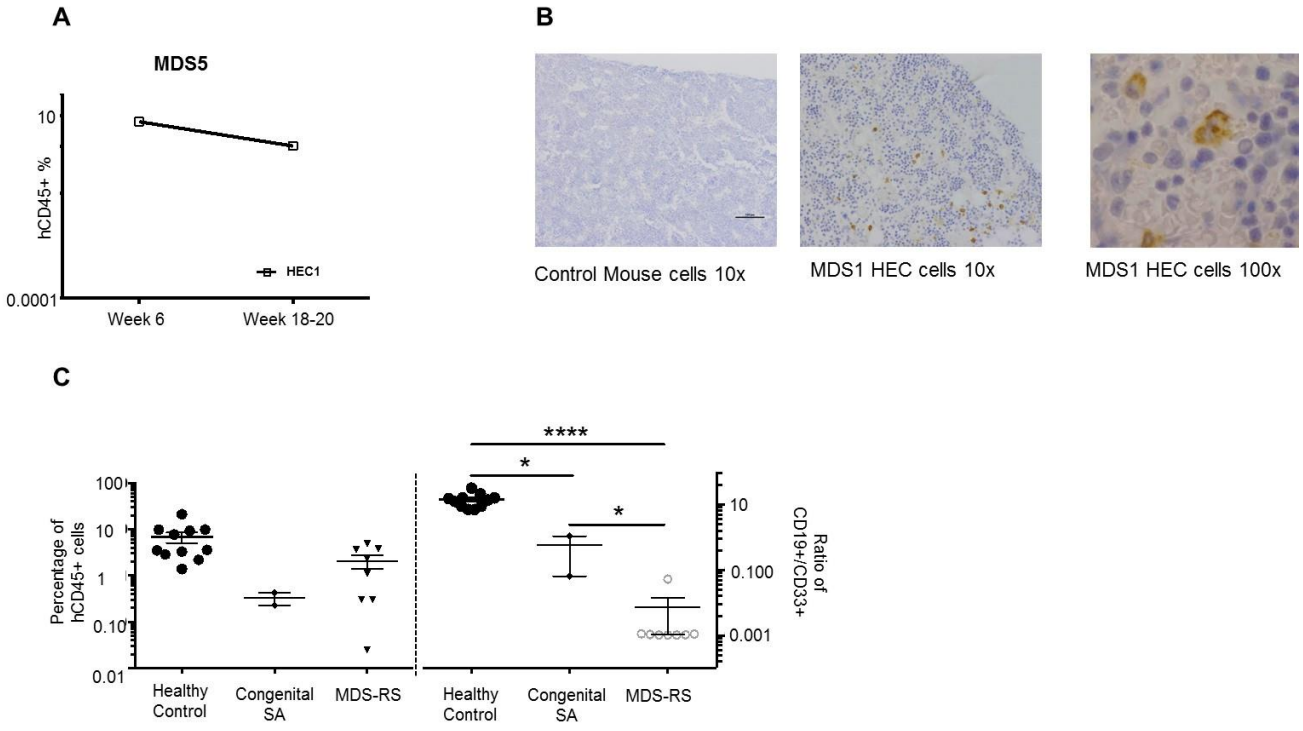


Figure 52: Engraftment of hematopoietic cells in the bone marrow of NOD/SCID/IL2rγ−/− (NSG) mice.

*(*A) Percentage of human CD45$^+$ cells in the bone marrow of NSG mice at 6 and 18-20 weeks after transplantation for MDS 5 patient. (B) Immunohistochemistry staining for human CD45$^+$ cells in the Xenograft bone marrow sample. Femurs were collected from engrafted and non-engrafted mice, and stained for human CD45 (in brown) and counterstained with Haematoxylin (in blue). (C) Percentages of human hematopoietic cells and ratio of CD19$^+$/CD33$^+$ isolated from the bone marrow of mice engrafted with either MDS-RS (n=4 patients and transplanted in a total 8 mice) or healthy controls CD34$^+$ cells (n=3 healthy donors and transplanted in a total of 11 mice) or congenital sideroblastic anaemia patient (n=1 patient transplanted in 2 mice). Y-axis (left) represents the percentage of the human CD45$^+$ cells present in the total mouse bone marrow. Y-axis (right) represents the ratio of the human CD19$^+$ vs human CD33$^+$ cells within the human CD45$^+$ cells recovered following xenotransplant. MDS xenografts showed a significant skewing toward myeloid lineage. * p-value <0.05, **** p-value <0.0001.

In order to determine the nature of the engrafted human CD45$^+$CD33$^+$ cells recovered from mice, mutational analysis was performed on HECs (human cells retrieved after the xenotransplantation; Table 35) using *SF3B1* mutation specific primers and subsequent sequencing through MiSeq platform (See methods section 2.2.6 for more details). Interestingly, the genotypic characteristics of patients' primary CD34$^+$ bone

marrow cells were maintained in HEC, as demonstrated by the presence of *SF3B1* mutations. *SF3B1* MAB remained largely unchanged between the primary $CD34^+$ cells (≈43%), and post-transplant HEC (≈46%) in 3/4 cases (Figure 51C). This was irrespective of the type of *SF3B1* mutations (Lys700Glu, n=2; His662Gln, n=2). In one RARS patient (MDS3), *SF3B1* (Lys700GLu) MAB increased from 28% in primary $CD34^+$ cells to 43% (average MAB between 3 mice) in xenografted HECs, therefore, suggesting a possible clonal advantage for the malignant mutant clone(s) over cohabitating normal HSCs in this case.

Table 35: Table showing a complete list of somatic mutations in 4 MDS-RS patients studied for *in-vivo* and *in-vitro* experiments.

Mutations present in paired primary $CD34^+$ cells, HEC samples and LTC samples. All mutations were confirmed independently by sequencing from 3 independent PCR reactions from all experiments. HEC- Human engrafted cells retrieved from mice. LTC- Long-term culture, N/A- Not applicable, MAB- Mutant Allele Burden, BM-bone marrow.

Experiment	Gene	Gene ID	Chromosome	Chromosome Position	Amino Acid Change	Primary BM MAB	HEC1 MAB	HEC2 MAB	HEC3 MAB	LTC MAB
MDS1	SF3B1	NM_012433	2	198266834	Lys700Glu	43%	46%	46%	50%	40%
MDS1	JAK2	NM_004972	9	5073770	Val617Phe	36%	17%	10%	17%	26%
MDS1	DNMT3A	NM_153759	2	25470588	Val107Leu	4%	10%	9%	12%	16%
MDS1	FBLIM1	NM_017556	1	16091611	Ala45Thr	27%	11%	6%	8%	17%
MDS1	HIST1H2AC	NM_003512	6	26124820	Lys120Asn	31%	14%	10%	14%	17%
MDS1	HSPA12A	NM_025015	10	118434314	Gly669Glu	2%	7%	7%	6%	9%
MDS1	RETNLB	NM_032579	3	108475355	Gly70Arg	28%	15%	7%	12%	15%
MDS1	SMC6	NM_024624	2	17897511	Lys456Met	2%	18%	16%	16%	9%
MDS1	SP3	NM_001017371	2	174777831	Asp598Asn	4%	19%	12%	18%	10%
MDS1	TH	NM_000360	11	2189733	Ala159Thr	15%	5%	4%	5%	8%
MDS2	SF3B1	NM_012433	2	198267371	His662Gln	42%	38%	N/A	N/A	39%
MDS2	DNMT3A	NM_175629	2	25467133	Trp581Ser	41%	39%	N/A	N/A	44%
MDS2	NAGA	NM_000262	22	42463872	Tyr74Phe	35%	44%	N/A	N/A	46%
MDS2	NMNAT2	NM_170706	1	183230405	Val224Met	41%	40%	N/A	N/A	45%
MDS2	NR0B2	NM_021969	1	27238545	Gly189Arg	29%	31%	N/A	N/A	25%
MDS2	SPATA13	NM_001166271	13	24797870	Arg268Gln	41%	44%	N/A	N/A	47%
MDS2	TLL2	NM_012465	10	98145885	Val647Ala	7%	1%	N/A	N/A	7%
MDS2	G3BP1	NM_005754	5	151175137	SS Exon 6 G>C	40%	45%	N/A	N/A	44%
MDS2	ABCB11	NM_003742	2	169833167	Gly410Ser	38%	32%	N/A	N/A	27%
MDS2	CCDC17	NM_001190182	1	46088498	Leu213Pro	17%	17%	N/A	N/A	24%
MDS3	SF3B1	NM_012433	2	198267371	His662Gln	28%	45%	36%	46%	41%
MDS3	ME1	NM_002395	6	84108234	Tyr72His	31%	37%	48%	53%	40%
MDS3	CAMTA1	NM_015215	1	7731086	Cys923Phe	33%	50%	53%	50%	39%
MDS3	PTPDC1	NM_001253829	9	96860488	Pro357Leu	21%	39%	35%	44%	35%
MDS4	SF3B1	NM_012433	2	198266834	Lys700Glu	40%	47%	N/A	N/A	N/A
MDS4	MYH14	NM_001145809	19	50728875	Pro251Ala	44%	46%	N/A	N/A	N/A

i

5.2.6 *SF3B1* mutant Haematopoietic stem cells engraft in NSG mice

Next, we sought to determine which progenitor cells within the $CD34^+$ hematopoietic compartment were able to engraft in NSG mice. To address this, bone marrow cells from one MDS-RS patients (MDS 14, S*F3B1* His662Gln mutation) were prepared and sorted on a BD FACS Aria SORP (San Jose, CA) operating in 4-way purity sort mode, and collected into 1.5 ml microfuge tubes (see methods section 2.2.8 for more details). Bone marrow cells were gated within the FACS plots to isolate HSC, MPP, MLP, CMP and GMP. After sorting, the purity of MPP, MLP, CMP and GMP were checked, and a sorting purity of ≥97% was obtained. Cell number obtained for each compartment for all the patients are mentioned in Table 36. Subsequently, all the cell fractions were injected separately into NSG mice. After 12 weeks, mice were sacrificed and bone marrow cells were analysed. Interestingly, human cell engraftment (HEC) was observed in only one mouse that received *SF3B1* mutated HSCs, which is similar to the recent data from Woll & colleagues (Woll et al., 2014) showing that MDS initiating cells (MDS-ICs) are restricted to the HSC compartment in $5q^-$ MDS patients (Table 36 and Figure 53). Mutational analysis targeting *SF3B1* gene was performed on $CD34^+$ cells and human engrafted cells obtained from mice. Our sequencing data confirmed that the engrafted cells carried the same *SF3B1* H662Q mutation with the similar MAB as compared to the day 0 $CD34^+$ cells from the patient (Figure 53). Therefore, our data suggest that $CD49^+$ HSCs with *SF3B1* mutations maintain their engraftment capacity. However, *SF3B1* mutations in this case did not increase the engraftment potential and/or induce the self-renewal capacity to the other cell types.

Table 36: Table showing the engraftment of FACS isolated haematopoietic cell fractions injected into NSG mice.

Progenitor cells from one MDS-RS patient were isolated and injected separately into NSG mice. 12 weeks later mice were sacrificed and bone marrow was analysed. Only the rare HSC faction gave rise to myeloid restricted engraftment indicating the nature of the MDS-SC in this rare HSC fraction.

		SF3B1 mutation CD34+ cells	Mouse Model	Cells Injected	Engrafted Mice	hCD33% Cells hCD19% Cells (12 weeks)	*SF3B1* MAB in CD34+ Cells (Day0)	*SF3B1* MAB in HEC (12 weeks)
Lin$^-$ CD45$^+$CD34$^+$CD38$^-$	**HSC** **CD45RA$^-$CD90$^+$CD49f$^+$**	H662Q	NSG	1629	1/1	100% 0.00%	46%	46%
	MPP **CD45RA$^-$CD90$^-$CD49f$^-$**	H662Q	NSG	29944	0/1	N/A	N/A	N/A
Lin$^-$ CD45$^+$CD34$^+$CD38$^+$	**MLP** **CD45RA$^+$CD90$^-$**	H662Q	NSG	51	0/1	N/A	N/A	N/A
	CMP **CD45RA$^-$CD135$^+$**	H662Q	NSG	16231	0/1	N/A	N/A	N/A
	GMP **CD45RA$^+$CD135$^+$**	H662Q	NSG	28217	0/1	N/A	N/A	N/A

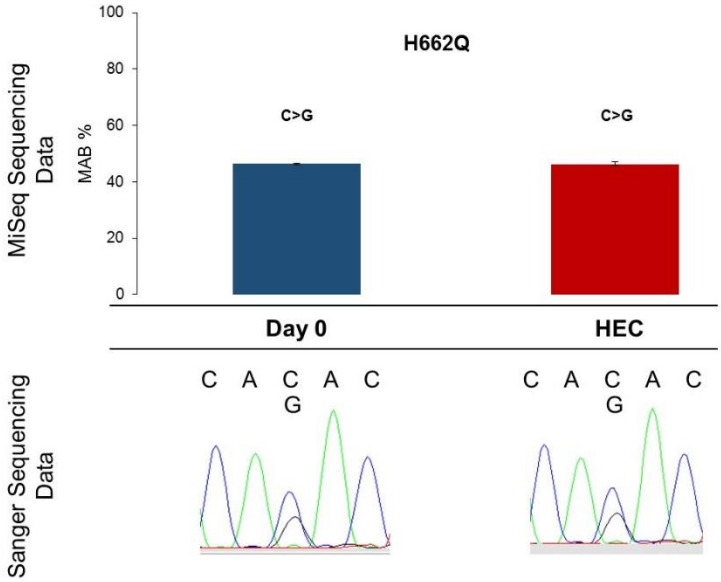


Figure 53: *SF3B1* mutation is maintained only in mice transplanted with HSCs.
Targeted mutational analysis shows the presence of concordant *SF3B1* mutations in primary $CD34^+$ bone marrow sample (Blue) and xenograft (Red) using MiSeq sequencer (top). Three independent PCR/sequencing experiments were performed to confirm/determine the mutant allele burden throughout the experiments using MiSeq sequencer. Sanger sequencing (bottom) traces confirm the mutation in primary $CD34^+$ bone marrow sample and xenograft. Error bars were generated from the PCR technical replicates. MAB- Mutant allele burden.

5.2.7 RARS Xenograft mice exhibits subclonal evolution of *SF3B1* mutant clones

To study the clonal architecture of the *SF3B1* mutant cells, WES and/or targeted mutational analysis was performed on HECs. This strategy enabled us to reliably trace *SF3B1* and known additional patient specific mutations as well as detect any new gene mutations, which may have 'evolved' during 18-20 weeks in NSG mice.

Figure 54: Mutational architecture and sub-clonal evolution of *SF3B1* mutated MDS-RS bone marrow cells.

Panel A, B, C and D shows the gene mutations with their MABs in primary CD34$^+$ bone marrow, HEC cells, LTC and pooled-CFC samples. Table on the right side of each panel represents the mutation status of the respective genes in the HSC, MPP, GMP and MEP cell populations for each patient. Bars representing the MABs for HEC samples is the average between ≥1 mouse samples (where applicable). Three independent PCRs were performed to confirm/determine the mutant allele burden throughout the experiments. (+) represents the increased MAB while (-) represent the decrease in the MAB in the HEC, LTC and CFC samples. *- represents the significant changes in the MABs (t-test P <0.05). P-mutant allele burden for gene mutations was same as the primary CD34$^+$ cells. PL- mutant allele burden for gene mutations was ≤ 2-fold as compared to the primary CD34$^+$ cells. PH- mutant allele burden for gene mutations was ≥2-fold as compared to the primary CD34$^+$ cells. ND- Not detected or below background noise level. MAB-Mutant allele burden. HEC-Human engrafted cells. LTC-Long-term culture. CFC- Colony-forming cells. BM- bone marrow. Mut- mutation.

Interestingly, patient-specific somatic mutations were generally maintained in HEC. However, changes in the MAB for some of these mutations were observed in 2/4 cases, indicating a subclonal population shift (Figure 54 and Table 35). In MDS1, *SF3B1* (Lys700Glu) MAB in primary CD34$^+$ cells and HEC remained almost identical at 43% and 47%, respectively (Figure 54). In contrast, *JAK2* (Val617Phe) MAB decreased from 36% in primary CD34$^+$ cells to 15% in HECs (average MAB for 3 mice). Likewise, a similar change in MAB for some additional gene mutations was also observed. Furthermore, HEC samples showed the emergence of divergent subclones with mutations in *DNMT3A* (≈10%) and other genes including *SMC6* (≈17%), *SP3* (≈16%) and *HSPA12A* (≈7%). These mutations were not initially detected in primary CD34$^+$ cells by WES. However, a follow-up high-depth targeted mutational analysis confirmed the presence of these mutations at a low level (<5%) in the primary CD34$^+$ cells (Figure 54). Similar results were observed for all gene mutations in pooled-CFCs and LTCs (Figure 54).

In MDS2, *SF3B1* (H662Q) mutation and additional gene mutations including *DNMT3A* MAB remained similar in the primary CD34$^+$ cells, HECs and LTCs as well as in pooled-CFCs (Figure 54). However, the *TLL2* gene MAB decreased from 7% in the primary CD34$^+$ cells to 1% in human engrafted mice (Figure 54).

In MDS3, where the *SF3B1* (His662Gln) mutation was detected at 28% in the primary CD34$^+$ cells, an increase in the MAB was observed in HEC (≈43%, average MAB between 3 mice). A similar increase was observed for additional gene mutations in HEC samples, LTC derived cells and pooled-CFC as compared to the primary CD34$^+$ cells (Figure 54), suggestive of the presence of a single *SF3B1* clone with multiple gene mutations.

In MDS4, only one aberration of *MYH14* gene was detected in addition to *SF3B1* where both had an MAB that was similar in primary CD34$^+$ cells (≈44%) and HEC sample (≈46%) (Figure 54), suggesting again the presence of a single clone carrying both *SF3B1* and *MYH14* mutations and as such no potential subclonal evolution in transplantation experiments.

Together this data demonstrates that *SF3B1* mutant clones are maintained in NSG mice and that subclonal fluctuation are detected in certain patients. Of note, all the

mutations present in the NSG mice were traced back to the patients' primary $CD34^+$ cells and no new mutations were detected or arose in the human engrafted stem cells from MDS-RS patients in NSG mice.

5.2.8 Patient Specific mutations are present in haematopoietic progenitors

WES data from 3 of the patients (MDS1, MDS2 and MDS4) identified additional patient specific mutations in $CD34^+$ cells. We sought to determine whether these patient specific mutations originate from the same HSC pool as *SF3B1* or whether they occur downstream of the stem cell differentiation in the progenitors. Mutational analysis demonstrated that the majority of the patient specific mutations including clinically important myeloid specific genes such as *DNMT3A* and *JAK2* were present in the rare HSC population as well as in the mature myeloid progenitors i.e. MPPs, GMPs and MEPs (Figure 54). However, *NORB2* and *TLL2* gene mutations for MDS2 were not detected in the HSCs but were present in the MPPs, GMPs and MEPs, suggesting that these mutations originate from a more mature progenitor compartment or are below the level of detection in HSC compartment. Furthermore*, SMC6*, *SP3*, *FLIMB1* and *TH* gene mutations were observed in the HSCs and MPP populations, and not in mature progenitors obtained from MDS1. This could be due to the fact that these mutations are present as minor clones (<10% MAB) in the HSC and MPP progenitor compartments and are as such diluted further in mature progenitors. Furthermore, we and others have shown that *SF3B1* mutations are not detected in T- or B-cells in MDS-RS patients. Taken together, this data suggests that *SF3B1* mutations are clonally propagated by HSCs to their mature myeloid progenitors but not to the lymphoid progenitors in MDS-RS patients.

5.2.9 Single cell clonal analysis reveals sub-clonal mutational architecture in MDS patients with RS

Previous studies have described mutational spectrum of *SF3B1* mutated patient bone marrow cells (Haferlach et al., 2014; Mian et al., 2013; Papaemmanuil et al., 2011; Yoshida et al., 2011b). However, these studies have inferred the clonal architecture in the total bone marrow samples through the identification of clusters of coexisting mutations with similar MABs and this does not necessarily recapitulate the actual subclonal architecture of the patients' bone marrow compartment. Therefore, to gain

insight into the clonal architecture of co-existing mutations in *SF3B1* mutant MDS-RS patients, we performed a single cell clonogenic assay using $CD34^+$ cells from 7 patients (MDS1, MDS2, MDS3, MDS4, MDS6 and MDS142).

After a 14 day CFC culture, individual colonies were picked, harvested and washed twice with PBS. Individual colonies (Table 33) were collected and stored in PBS. Following on, whole-genome amplification was performed on CFC colonies (WGA; See methods section 2.2.13 for more details). WGA amplified colonies were subjected targeted mutational analysis of the patient specific mutations (mutational profile obtained from WES, Table 35) using MiSeq platform (See methods section 2.2.6 for more details). Gene mutation for each single colony was called when ≥30% of the sequencing reads were carrying a mutant allele.

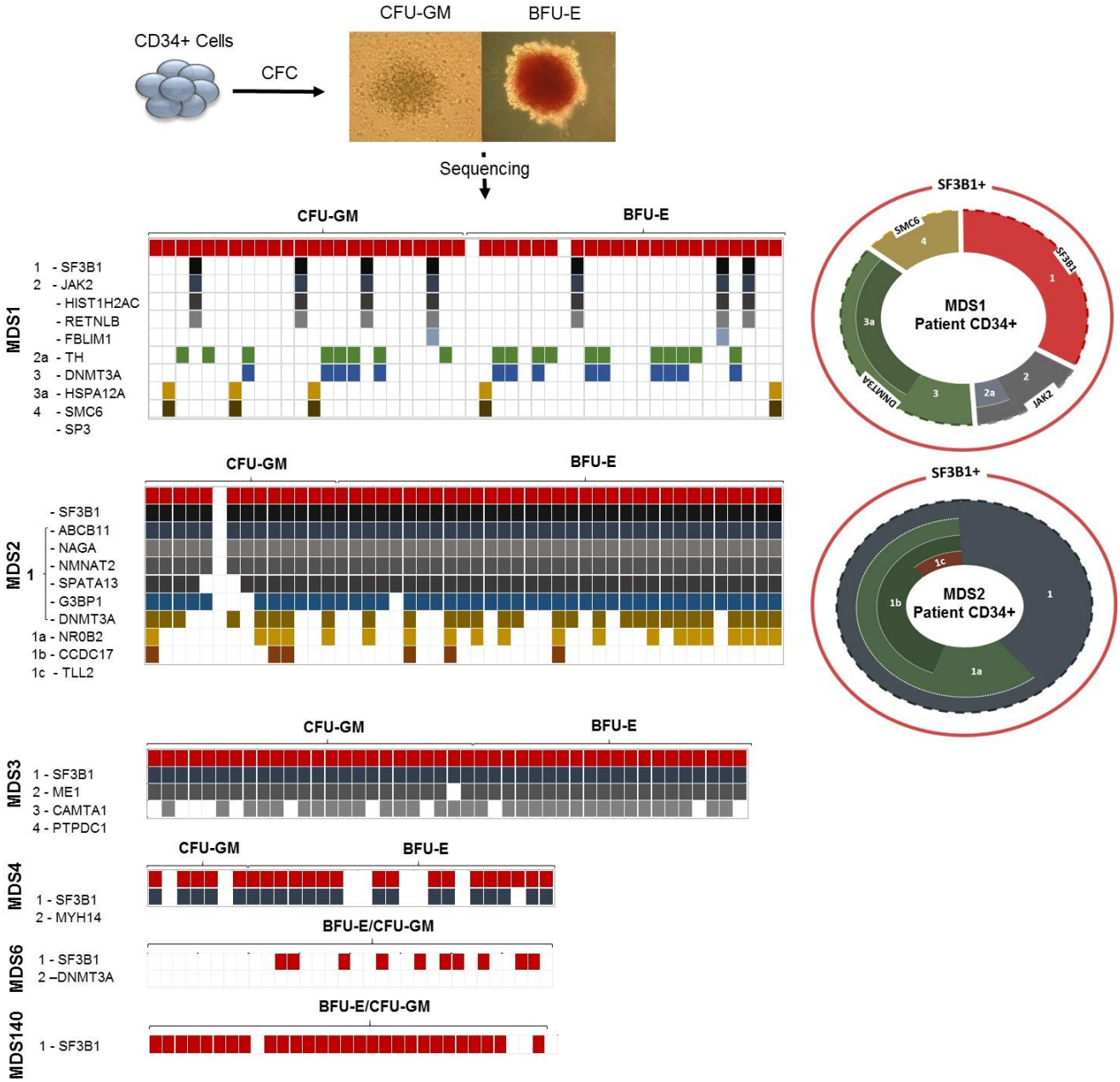


Figure 55: *SF3B1* mutations in single cell colonies.

Mutational analysis was performed on BFU-E and CFU-GM colonies derived from primary CD34^{+} patient cells (MDS1, MDS2, MDS3 and MDS4). Each column represents an individual CFC colony. Circle (MDS1 and MDS2) on the right represent the clonality observed in the CD34^{+} cells in MDS1 and MDS2. WT- Wild Type. Coloured boxes represent a mutant colony while as white box represent a wildtype colony.

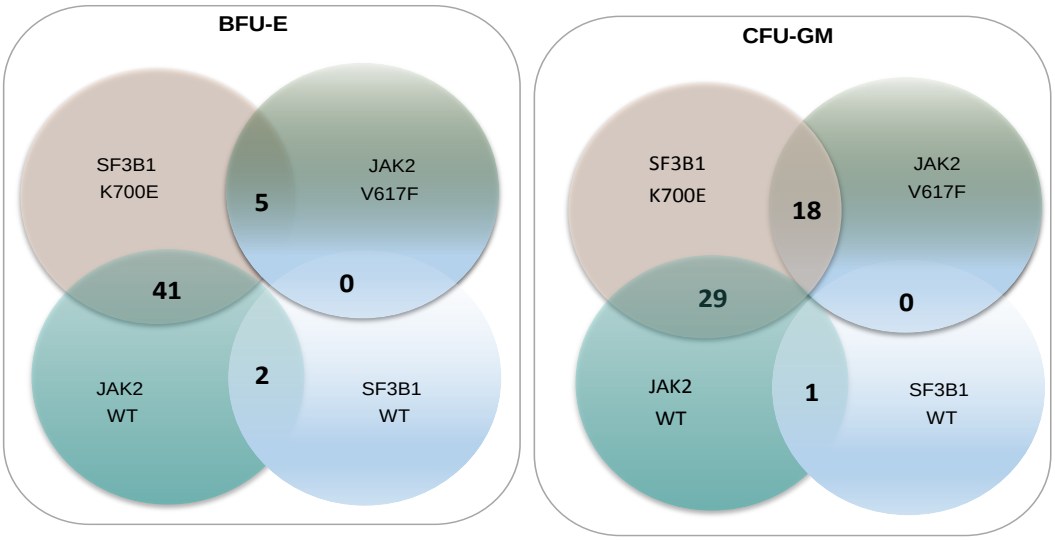


Figure 56: *In-vitro* analysis of *SF3B1* mutant MDS1 patient CD34$^+$ cells.
Mutational analysis performed on BFU-E and CFU-GM colonies derived from primary CD34$^+$ patient cells from MDS1. 96 individual colonies were screened for *JAK2* and *SF3B1* mutations. *SF3B1* mutations are seen equally in BFU-E and CFU-GM colonies.

Initial mutational analysis performed on individual CFC colonies from MDS1 (*SF3B1* and *JAK2* screened in 96 colonies; other patient specific mutations screened in 48/96 colonies), MDS2 (47 colonies), MDS3 (46 colonies) and MDS4 (30 colonies) using gene specific primers targeting *SF3B1* mutations revealed that 97%, 98%, 96% and 77% of the screened colonies carried *SF3B1* mutant gene respectively (Figure 55 and Figure 56). However, further analysis of the 'patient specific' gene mutations in these cases revealed the complex heterogeneity within the *SF3B1* mutated colonies. For example, MDS 2 had a marked subclonal heterogeneity with 7 *SF3B1* subclones containing 11 mutated genes with majority (46/47 colonies) of them related via a branching ancestral tree (Figure 54). Notably, driver mutations such as *DNMT3A* were present in 44/47 colonies, suggesting that it may have been acquired at a later time point during disease progression. Furthermore, 48/96 colonies analysed from MDS1 demonstrated 6 different *SF3B1* subclones; only 15/48 colonies had *SF3B1* mutations alone (Figure 55 and Figure 56). Six *SF3B1* subclones were mutually exclusive especially subclones with known driver mutant genes such as *JAK2* and *DNMT3A*, therefore representing a branching multi-clonal evolution (Figure 56). In contrast, the simplest of the genetic architectures was present in MDS142, MDS6, MDS3 and

MDS4 where only one, two and three mutant *SF3B1* subclones were observed, respectively (Figure 56).

It is noteworthy that CFC individual colonies with wild-type *SF3B1* (MDS1, n =2; MDS2, n=1; MDS3, n=2, MDS4, n=7, MDS6, n= 20) had no additional gene mutations amongst those screened (Figure 55 and Figure 56). This provide further evidence that *SF3B1* mutations are present in the majority of the $CD34^+$ cells and, importantly, precede other driver mutations such as *DNMT3A* and *JAK2* as evidenced here. This implies that *SF3B1* mutation is the 'founder mutation' in these MDS-RS patients.

5.2.10 Sequential acquisition of genetic lesions in *SF3B1* mutant clones leads to disease transformation

MDS2 patient transformed to AML 29 months after the first MDS sample was taken (MDS2, Time point 1). Initially, FACS analysis was performed to examine the phenotypic changes that could have occurred in the bone marrow compartment leading to AML (Figure 57A). The frequency of $CD34^+$ cells in bone marrow MNCs at AML stage (at 29 months) was significantly higher as compared to the MDS stage (MDS stage time point 1, $CD34^+$ = 1.54% of total MNCs; AML stage $CD34^+$ = 35.3% of total MNCs). A remarkable shift of cell populations was observed within the $CD34^+CD38^-$ compartment (HSCs, MPPs and MLPs) as well as $CD34^+CD38^+$ more mature progenitors (CMPs, GMPs and MEPs) at the AML stage as compared to MDS stage of the disease (Figure 57A). By contrast to MDS stage, the majority (>99%) of the cells observed at the AML stage within the $CD34^+CD38^-$ compartment were MLPs. Similarly, >99% of the cells detected in the $CD34^+CD38^+$ compartment at the AML stage were GMPs.

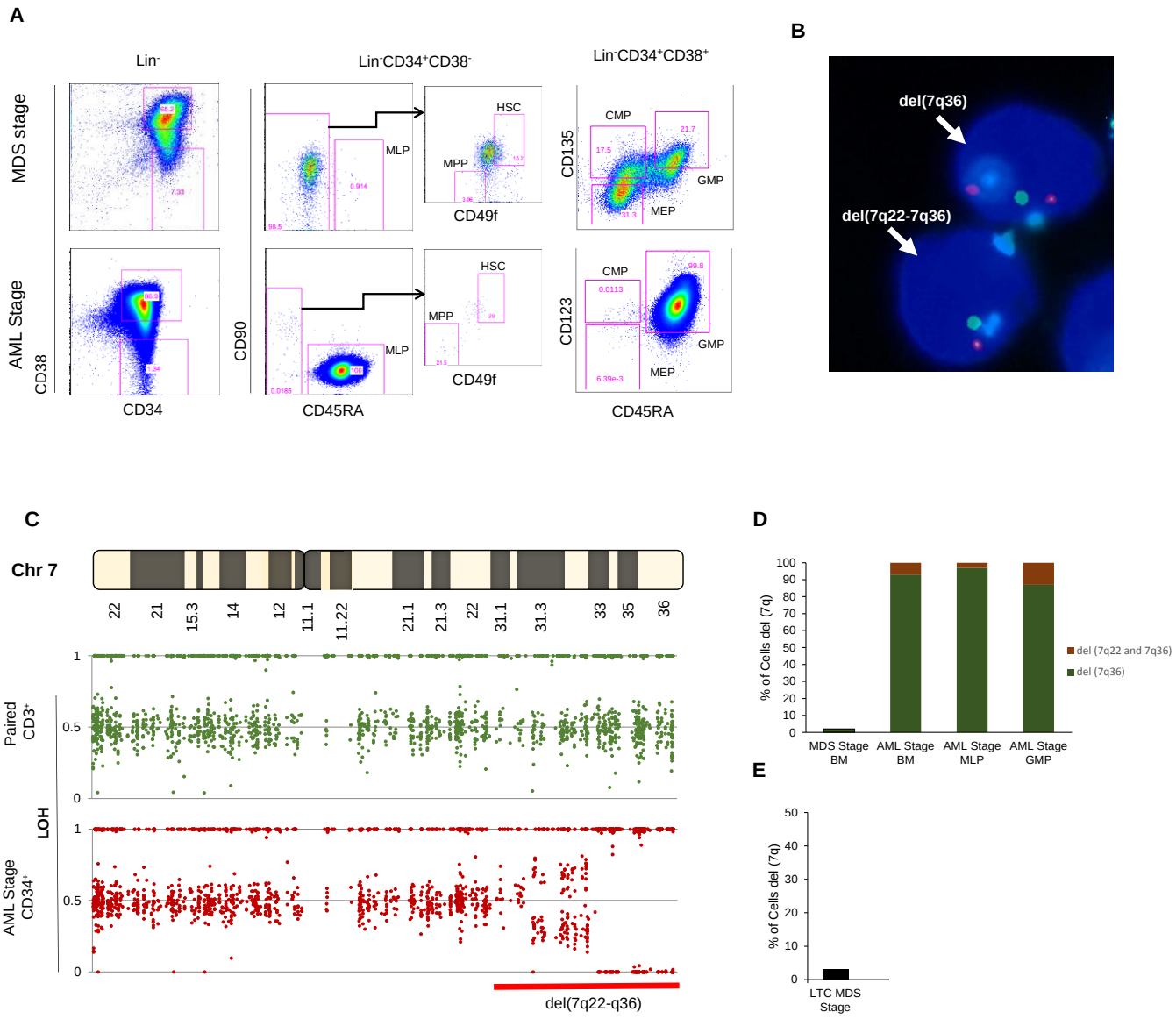


Figure 57: Clonal evolution from MDS to AML in RCMD-RS patient with *SF3B1* mutation. (A) FACS profiles of Stem cell progenitors in BM MNCs in serial samples at MDS stage and AML stage of the disease. (B) Interphase FISH for tracking the del(7q) aberration in cells obtained at AML stage of the disease. Two chromosome 7 aberrations del(7q36) and del(7q22-7q36) were detected as two separate clones. Two individual cells are shown in the figure with one harbouring del(7q36) represented by green probe and the other cell has del(7q22-7q36) as represented by probes orange/green. (C) SNV profile derived from whole-exome sequencing. The profile depicts a heterozygous deletion (horizontal red line) of chromosome 7q in patient $CD34^+$ bone marrow cells (bottom) and paired control $CD3^+$ cells (top). (D) Bar chart (FISH analysis) showing percentage of $CD34^+$ cells (MDS stage), $CD34^+$ cells (AML stage), MPL (AML stage), GMP (AML stage) with del(7q) aberration.

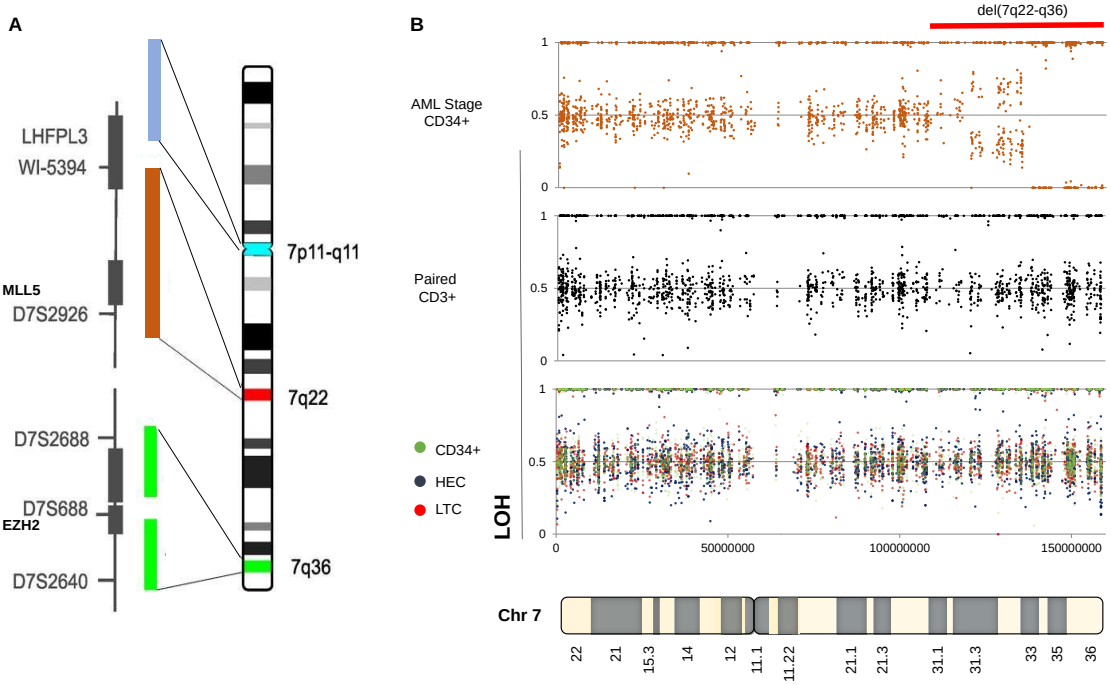


Figure 58: FISH and SNP (whole-exome) profiling for MDS2.

(A) Detailed map of locus specific probe targeting q arm of chromosome 7. XL 7q22/7q36 probe detects a specific region at 7q22 (orange) including the MLL5 gene while as the green labelled probe hybridizes specifically to 7q36 (green) and includes the EZH2 gene. In addition, a blue (aqua) labelled probe which hybridizes to the centromeric region of chromosome 7 acts as a reference probe. Chr-Chromosome. (B) Single nucleotide variations (SNVs) profiles from whole-exome experiments based on $CD34^+$ cells (AML stage), paired constitutional DNA ($CD3^+$), $CD34^+$ (MDS stage time point1), Human engrafted cells (post engraftment, MDS stage time point 1) and LTC derived cells (MDS stage time point 1). SNVs were selected that were called germline or LOH in VarScan, due to presence at similar levels in both samples of a pair or significant allele burden change between the two respectively, had >20 read depth. LOH- Loss of heterozygosity.

We sought to investigate the contribution of *SF3B1* and additional genomic lesions (chromosomal aberrations; gene mutations, known 'driver' or 'passenger' mutations) were acquired at the AML stage. Metaphase cytogenetics at the AML stage revealed complex chromosomal changes (Table 32) including del(7q). In order to explore this further and trace back the del(7q) aberration, FISH analysis (Figure 58A-58B and Figure 58B) for chromosome 7 was performed on the whole bone marrow sample (AML stage), MLP (AML stage), GMP (AML stage) and whole bone marrow sample

(MDS stage, Time point 1). Two chromosome 7 aberrations were detected in the whole bone marrow sample taken at the AML stage of the disease with 93% (140/150 cells) having del (7q36) and 7% (10/150 cells) containing del(7q22-7q36). Similar results were observed in the MLPs (del7q36, 292/300 cells; del7q22-7q36, 8/300 cells) and GMPs (del7q36, 262/300 cells; del7q22-7q36, 38/300 cells) obtained at AML transformation (Figure 57B-57C). Only del(7q36) was traced back to the MDS stage of the disease, however, the frequency was significantly less i.e. 2% (12/600 cells) (Figure 57D-57E). FISH analysis couldn't be performed on the phenotypically defined HSCs, MPPs, CMPs and MEPs due to the low frequency of these progenitors in the bone marrow at the AML stage. It is noteworthy that FISH analysis on the LTC derived cells (MDS stage Time point 1) showed low level del (7q36) clone (3%, 8/300 cells; Figure 57D), therefore suggesting that the major clone containing del (7q36) is an early lesion acquired in the *SF3B1* mutant rare HSCs at an MDS stage of the disease. However, SNV profile through WES on LTC showed no change in the LOH at chromosome 7 and similar results were observed for $CD34^+$ cells (MDS stage), HEC (MDS stage) and $CD3^+$ cells (Figure 58) which could be due to the detection limitations of WES.

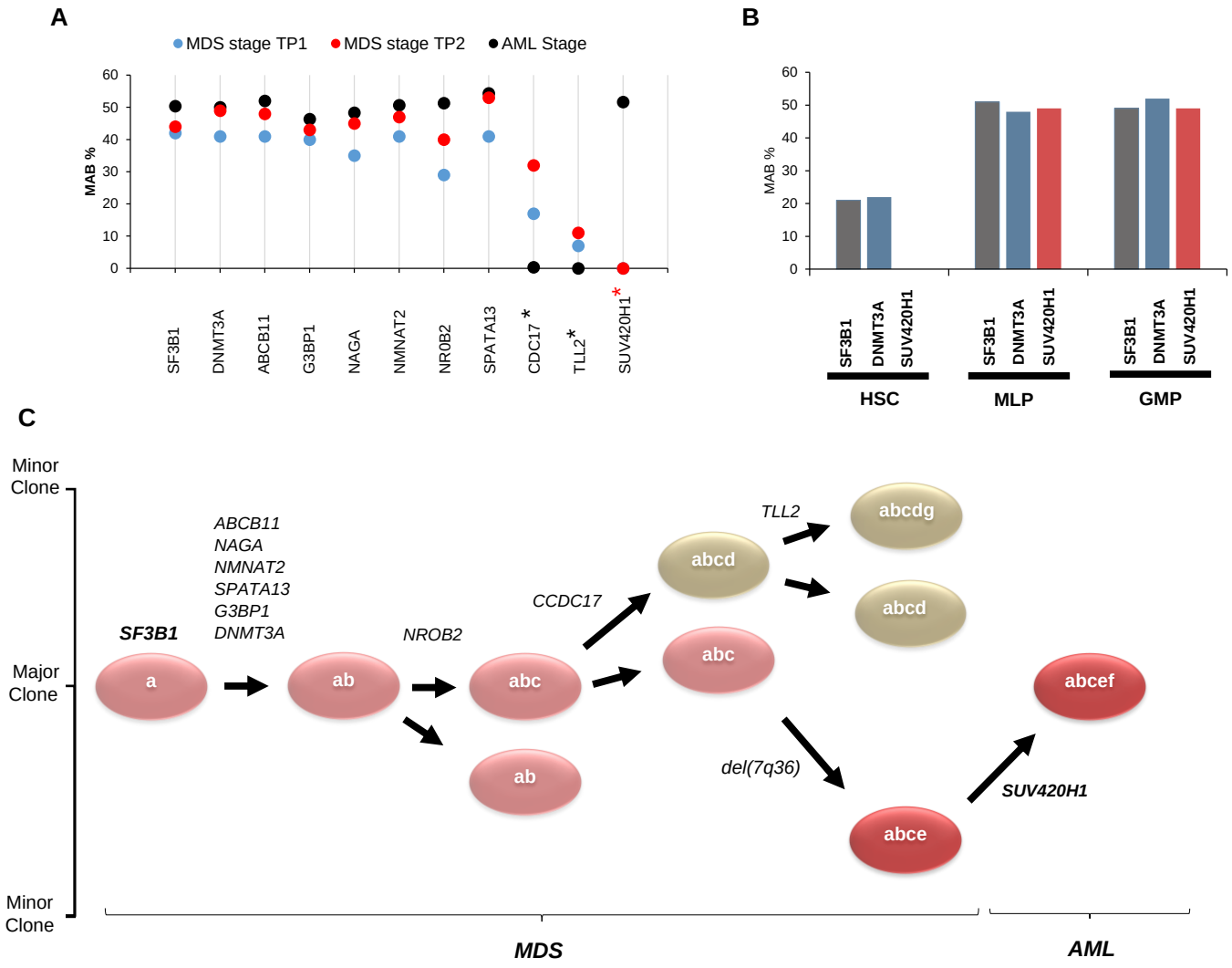


Figure 59: Clonal evolution from MDS to AML in RCMD-RS patient with *SF3B1* mutation. (A) Mutational status of the gene mutations in sequential bone marrow samples (MDS stage-time point 1, MDS stage-time point 2 and AML stage). Coloured circles represent the mutant allele burdens for each screened gene mutation. Black *- represent absence of gene mutations in AML stage sample. Red*- represents acquisition of additional mutation at the AML stage. (B) Mutational analysis of *SF3B1*, *DNMT3A* and *SUV420H1* in HSCs, MLPs and GMPs from AML stage of the disease. Sequencing depth for HSCs was >10000 reads. (C)Proposed sequential acquisition of genetic lesions based on analysis of sequential samples, LTC derived cells, and single cell clonogenic assays.

WES of the whole bone marrow sample (AML stage) showed that *SF3B1* and other additional gene mutations including *DNMT3A* were maintained, however, *CCDC17* and *TLL2* were not detected (Figure 59A). Further analysis revealed additional patient specific mutations, including *SUV420H1* (Leu645Phe) novel mutation (Figure 59A), which is a histone methyltransferase that specifically trimethylates 'Lys-20' of histone H4. Due to the functional importance of this gene and its known association with breast cancer (Fraga et al., 2005; Tryndyak et al., 2006), we followed this gene mutation and performed high depth targeted mutational screening on MDS stage samples (Time point 1 and Time point 2). Unlike del(7q), this mutation was not detected in any of the previous MDS samples (Figure 59B).

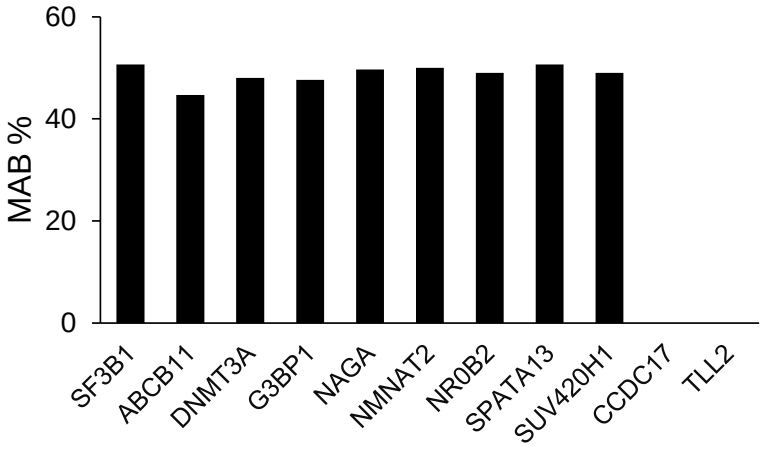


Figure 60: Mutational profile of sequential MLP sample for MDS2 obtained at AML stage of the disease. MAB- Mutant allele burden.

Furthermore, mutational analysis of the HSC, MLP and GMP compartments (AML stage) revealed that *SF3B1* and *DNMT3A* mutations were present in all the compartments. However, *SUV420H1* mutation was not detected in HSCs but originated from MLPs and propagated to GMPs. (Figure 59B and Figure 60). Taken together, our data suggests that *SF3B1* mutation precedes del(7q) aberration and is maintained throughout the disease evolution (Figure 59C).

5.2.11 RARS Xenograft model recapitulates the clonal changes occurring in patients bone marrow compartment

In order to ascertain whether the xenograft clonal changes observed, occur in the MDS-RS patient bone marrow, targeted mutational screening was performed on sequential samples (total nucleated cells) obtained from 2 patients (MDS1, ≈18 months after initial sampling with no progression and MDS2, ≈29 months after the first sampling at the AML transformation stage; Table 32).

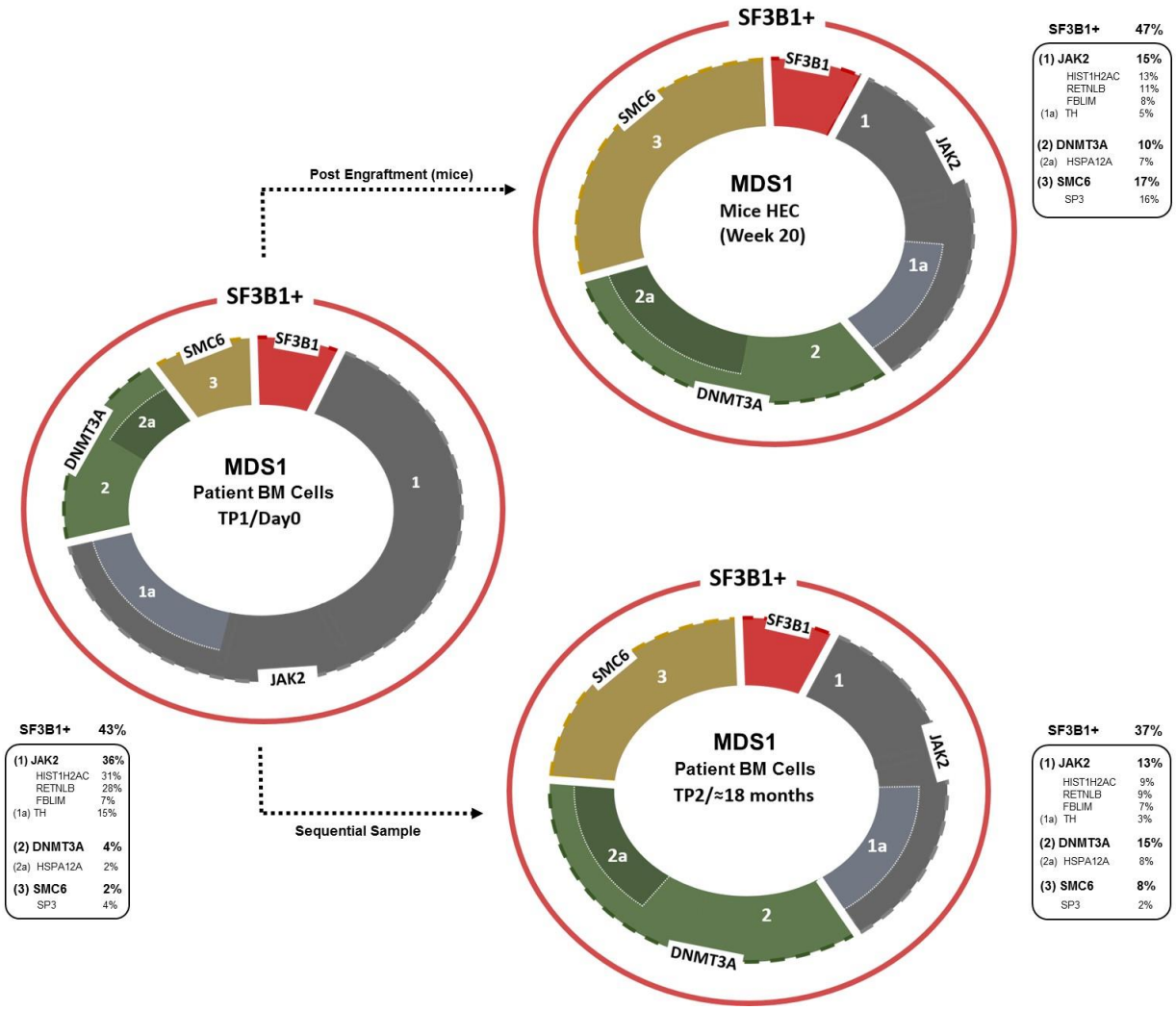


Figure 61: Xenograft recapitulates the clonal changes occurring in RARS patient bone marrow compartment.

Mutational analysis of sequential primary TNC (total nucleated cells) sample from 1 patient (MDS1). Primary patient sample was received at two different time points, MDS1 $CD34^{+}$ (TP1, first sample) and MDS1 TNC (TP2, second sample). Mutational analysis was also performed on human engrafted cells which were obtained from mice transplanted with patient sample received at time point 1. The length of the arc represents the mutant allele burdens. Gene MAB for human engrafted cells is the average between three animal experiments. Three independent PCR/sequencing experiments were performed to confirm/determine the mutant allele burden throughout the experiments. MAB- Mutant allele burden. TP-Time point, HEC-human engrafted cells, TNC- Total nucleated cells.

Mutational analysis of these sequential samples revealed that *SF3B1* mutations were largely unchanged between the two time points (Figure 61 and Figure 62) in both patients. By contrast, a change in the MAB of additional somatic mutations was observed in the latter sample obtained from MDS1 (Time point 2; Figure 61). In this case, MABs of some of the gene mutations including *JAK2* (Val617Phe, 36% to 13%) diminished (Figure 61), whereas *DNMT3A, HSP12A* and *SMC6* MABs increased from 4% to 15%, 2% to 8% and 2% to 8% in the patients sequential sample (Time point 2; Figure 61), respectively. Strikingly, xenografted cells (mice transplanted with sample at time point 1) from the same patient at week-20, demonstrated similar sub-clonal changes as observed in the sequential patient sample after ≈18 months (Time point 2, Figure 61). This data along with our clonogenic data suggest that 4 *SF3B1* subclones exist in this patient, however, "isolated-*SF3B1*" and "*SF3B1* plus *DNMT3A*" clone have a growth advantage over the rest of the clones, especially over "*SF3B1-JAK2*" in NSG mice as well as in human bone marrow environment (Time point 2).

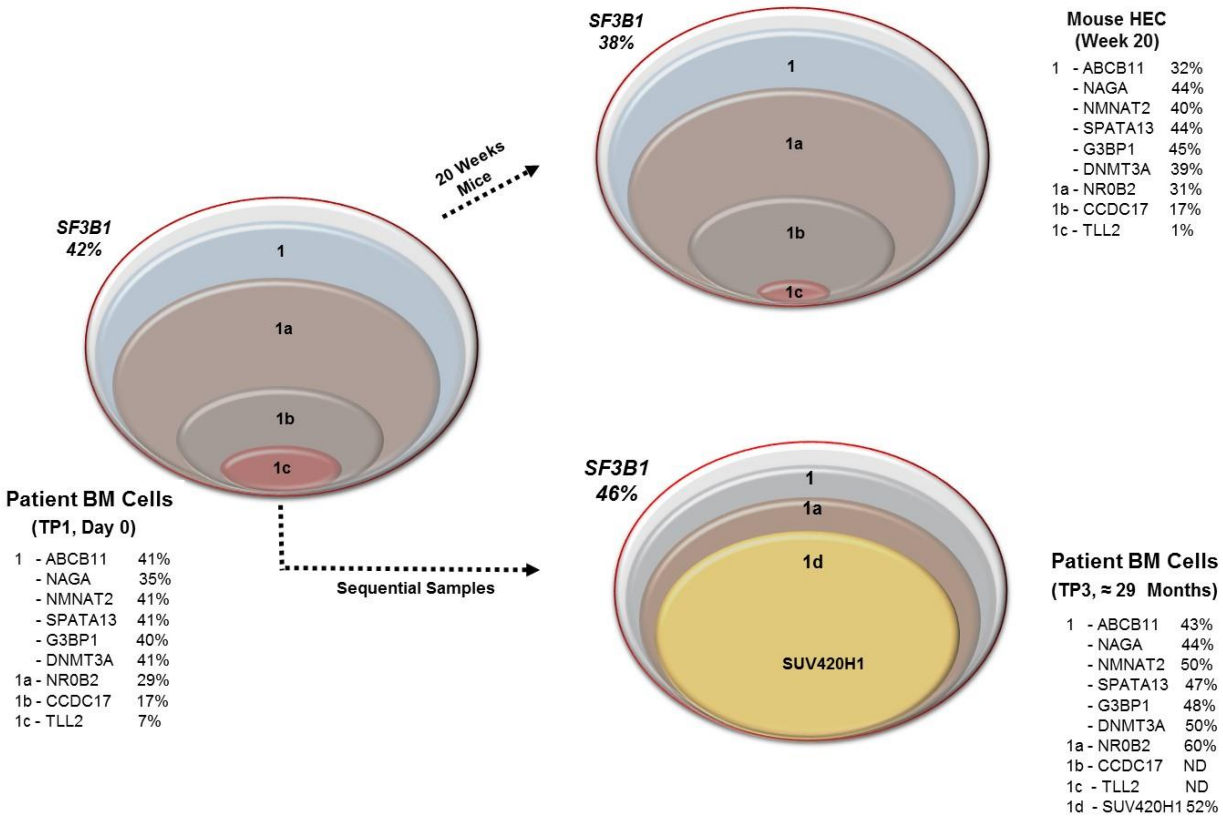


Figure 62: Xenograft recapitulates the clonal changes occurring in the patient bone marrow compartment.

Mutational analysis of sequential primary TNC sample from MDS 2. Primary samples were received at two different time points, MDS 2 $CD34^{+}$ (TP1, MDS stage) and MDS 2 TNC (TP3, AML Stage). Mutational analysis was also performed on human engrafted cells which were obtained from mice transplanted with patient sample taken at time point 1. Three independent PCR/sequencing experiments were performed to confirm/determine the mutant allele burden throughout the experiments. MAB- Mutant allele burden. TP-Time point, HEC-human engrafted cells, TNC- Total nucleated cells, ND- Not detected.

In MDS2, *CCDC17* and *TLL2* mutations were not detected in the sequential sample while the rest of the mutations were largely maintained. A similar trend was observed in our NSG mice where *TLL2* gene mutation diminished from 7% in the primary patient samples (MDS stage Time point 1) to 1% in the HEC sample (Figure 62). *SUV420H1* mutation which was detected in the AML stage after MDS2 transformed, was not detected in the HEC sample, perhaps reinforcing the fact that this mutation was

acquired at a later time point in the patients' primary bone marrow cells and was not part of the initial MDS clones injected in mice.

Altogether, this data suggest that xenograft model does recapitulate some of the clonal changes occurring in the patients' bone marrow compartment in one case while as in another case where AML transformation occurred, a partial recapitulation was observed most probably due to the stochastic acquisition of additional genomic lesion at this later stage of the disease (in this case *SUV420H1* mutation). Meanwhile, it is noteworthy that the 'incubation period' of the MDS bone marrow cells in NSG mice was only 20 weeks while patient (MDS2) transformed to AML 29 months after the first MDS sampling.

5.2.12 Proposed model for existence of a 'Master Switch' causing leukaemia transformation

SF3B1 mutation(s) in MDS patients with ring sideroblasts arise from $CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$HSCs and is the initiating event in disease pathogenesis. However, acquisition of additional 'potent' mutations such as *SUV420H1* in a patient with "SF3B1-del(7q)" subclone in the MLP compartment leads to the expansion of this 'lethal' MLP clone and hence transformation to AML (Figure 63).

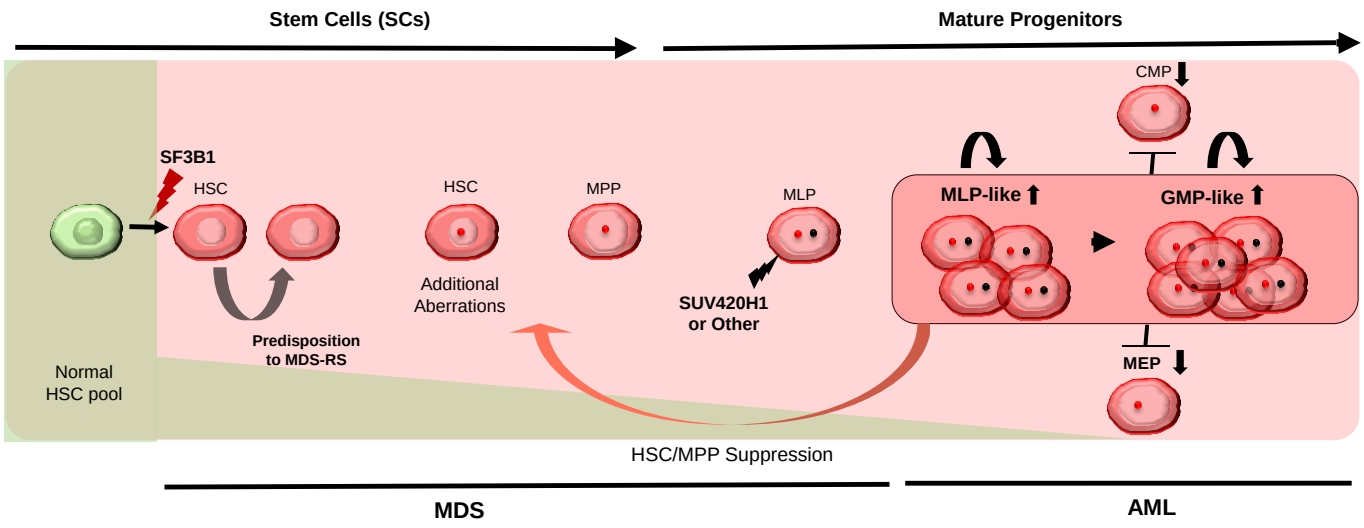


Figure 63: Proposed model for sequential acquisition of genetic lesions in MDS to AML transformation.

SF3B1 mutations are acquired in $CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}CD49f^{+}$ HSCs. As HSCs replenish downstream progenitors, mutations acquired in stem cells are also propagated to mature progenitors but without conferring self-renewal potential. However, as the mature progenitors such as MLPs harbouring a mutant *SF3B1* acquires a more "potent hit" such as *SUV420H1* mutation, they confer self-renewal potential and suppress the HSC to MPP transition.

5.3 Discussion

Next-generation sequencing technologies have greatly improved our understanding of the genetic landscape in cancer. Previous work in AML and MDS demonstrated the existence of clonal heterogeneity that evolves upon disease progression and/or relapse (Ding et al., 2012; Klco et al., 2014; Mardis et al., 2009; Walter et al., 2011; Welch et al., 2012). To date, various studies including ours have shown that *SF3B1* is the most frequently mutated gene in MDS with remarkably strong association with a specific disease phenotype i.e. the presence of ring sideroblasts (Mian et al., 2013; Papaemmanuil et al., 2011; Yoshida et al., 2011b) suggesting that this mutation is casually linked to event defining ring sideroblastic anaemia. However, it is noteworthy that studies so far have not determined the origins of *SF3B1* mutations within the hematopoietic compartment and the composition of the *SF3B1* subclones. Similarly, the functional heterogeneity of genetically defined *SF3B1* mutant subclones has not been delineated.

Here WES, capture-based targeted high-depth sequencing, long-term culture assay, xenotransplantation as well as single cell clonogenic assays were used to study the characteristics of *SF3B1* mutant hematopoietic cells. Mutational analysis of the *SF3B1* gene in hematopoietic stem/progenitor cells reveals that *SF3B1* mutations arise in phenotypically defined $CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$ stem cells therefore unveils the MDS-SCs existence in RARS that propagate mutations to their myeloid progenitors such as MPPs, GMPs and MEPs. This was also reflected in clonogenic assays where *SF3B1* mutations were detected either alone or in combination with other mutations in individual CFC colonies, therefore implying that *SF3B1* aberrations precede other gene mutations. Similar results have been described recently, in MDS patients with concurrent *SF3B1* and del(5q) aberrations, where *SF3B1* mutations were shown to be present in $CD34^{+}CD38^{-}CD90^{+}$ stem cells in these patients (Woll et al., 2014).

In our study, xenotransplantation of $CD34^{+}$ bone marrow from MDS samples in immuno-deficient mice demonstrated the persistent long-term myeloid restricted engraftment in the majority of cases. In contrast, NSG data from one congenital sideroblastic anaemia patient carrying an *ALAS2* mutation and other healthy donors

showed a multi-lineage engraftment with high percentage of lymphoid/myeloid cell ratio. It is noteworthy that all 4 patients in our study had only one common genomic aberrations i.e. *SF3B1*. In addition, 2/4 patients (MDS 2 and MDS3) had a lower mutation rate (2 to 4 mutations/exome) which is less as compared to that in AML (≈13), chronic lymphocytic leukaemia (≈12), and was significantly lower as compared to multiple myeloma (≈33), hence representing a much more homogenous clonal population.

Notably, the genotypic analysis of pre- and post-transplanted cells demonstrate that in most of the cases *SF3B1* mutations were maintained, however, in one case (MDS3), *SF3B1* mutant clone increased from 27% in primary $CD34^{+}$ cells to 47% in HEC samples. This suggests the presence of a mixed stem cell population (i.e. *SF3B1*-mutant/wild type), which when injected into immuno-deficient mice resulted in the expansion of *SF3B1* mutant clones preferentially.

Exclusive myeloid engraftment and the maintenance/increase in the *SF3B1* clone(s) in NSG xenografts suggests that a malignant HSC clone with *SF3B1* mutations either alone or in-conjunction with other “co-operating factors” has a growth advantage or provides a negative feedback over the cohabitating normal HSC.

Interestingly, in all mice that have detectable human engrafted cells, we could see a similar subclonal architecture compared to MDS cells injected. Nevertheless, in some cases, the engrafting *SF3B1* daughter subclones represented only a small fraction of the injected cells (<5%), implying that some *SF3B1* subclones have a cell-intrinsic advantage (due to better engraftment ability, enhanced proliferation potential, and/or additional factors) following xenotransplantation. Malignant clones with ‘isolated *SF3B1’* mutations are not always driven away by more ‘potent’ mutant subclones (such as *DNMT3A* and/or *JAK2*) in NSG mice as was reflected by its MAB which remained stable even though other subclones were fluctuating. Perhaps additional driver mutations when acquired by the *SF3B1* clones confer a subtle advantage to the new daughter subclone representative of early stages in disease progression as evidence for MDS1 (Time point 1 vs Time point 2). However, when these clones acquire more aggressive mutation(s) as it is the case here for MDS2, a more ‘aggressive’ mature progenitor clone with self-renewing capacity is generated, enabling the disease to evolve into AML. It is tempting to speculate that *SUV420H1*

mutations confers a growth advantage based on the evidence that Suv420h (referred to both Suv420h1 and Suv420h2) deficient MEFs and ES cells have been reported to have telomere elongation and epigenetic defects (Benetti et al., 2007). Further work needs to be done to study the role of *SUV420H1* mutation(s) in haematological diseases.

It is noteworthy that small subclone with *SF3B1* mutation harbouring del(7q36) becomes a predominant clone in the sequential AML sample (MDS2) but with the additional acquisition of del(7q22). Therefore, *SF3B1* mutation in isolation might only have a minor role in disease transformation or progression to AML. However, like del(5q) aberrations (Woll et al., 2014), it may instead be a strong driver mutation in MDS-RS and acquires additional mutations during the course of the disease progression.

Previous studies of functional heterogeneity in human leukaemi*as* have used NSG mice to characterize leukemia stem cells in order to decipher their engraftment potential, self-renewal capacity (Eppert et al., 2013; Goardon et al., 2011; Taussig et al., 2008; Taussig et al., 2010). Nevertheless, Tim Ley's group was the first to report functional heterogeneity of genetically defined subclones in AML (Klco et al., 2014). Their data suggested that xenotransplantation results in a dramatic decrease in the subclonal complexity of the AML samples and were not able to reliably predict disease relapse. However, the combination of our WES in xenotransplanted NSG mice and serial patient sampling enabled us to compare the clonal evolution observed in mice with that of patient bone marrow compartment in MDS1 and MDS2. The changes in the spectrum of *SF3B1* mutant clones observed in the NSG mice were largely reflected in the patient's sequential bone marrow cells where similar clonal changes were detected. Although the *SF3B1* clonal fluctuation observed in NSG mice for MDS1 was more remarkable as compared to MDS2.

This is the first study to demonstrate *SF3B1* mutations originate in rare $CD34^{+}CD38^{-}CD90^{+}CD49f^{+}$ HSCs and MDS-initiating cells are restricted to these rare HSCs fractions in MDS patients with ring sideroblasts. Although our data is limited to five patients, the presence of *SF3B1* mutations in $CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}CD49f^{+}$ HSCs provides new insight into the MDS stem cells. Furthermore our data

demonstrates that *SF3B1* mutations acts as an initiating event in MDS-RS disease, as has been suggested by previous studies (Papaemmanuil et al., 2011; Yoshida et al., 2011b). Moreover, MDS2 patient progression to AML illustrates the concept of a switch from “HSC disease” to a “Progenitor disease”. Here, we show that MDS-RS originates from the stem cells and at a later stage of the disease, *de novo* mutation(s) acquired by the progenitor cells confers the self-renewal capability to the MDS-progenitor cells that drives the leukemic transformation. This novel observation confirming the presence of MDS-SC at the AML stage provides new insight into the MDS/AML hierarchy. Our results also suggest that MDS-SC at the AML stage can constitute a pool of pre-leukemic cells providing a reservoir of cells that could be responsible for relapse. Although mutations such as *SF3B1* are not driving mutations at the AML stage, they provide a genuine target for curative therapies and give the opportunity to therapeutically target at the same time the AML blasts as well as pre-leukemic cells.

Chapter 6

General Discussion

Complete genomic characterization of MDS in-general has revealed a relatively restricted number of genes that are recurrently mutated and these genes encode proteins that are involved in a small number of key biological pathways. Among them, defects of pre-mRNA splicing genes, particularly *SF3B1* which is a key component of the spliceosome complex is mutated in up to 85% of the RARS/RCMD-RS patients as has been shown here and elsewhere (Papaemmanuil et al., 2011; Yoshida et al., 2011b). In addition, recurrent somatic mutations observed in a number of additional spliceosomal genes (such as *SRSF2*, *U2AF35*, *ZRSR2*) suggests that dysregulated RNA splicing may be driving the clonal dominance in at least 50% of the MDS patients. Homozygosity is the exception to the rule for most of the spliceosome mutations, especially *SF3B1,* and spliceosome mutations are mutually exclusive. Therefore, both these scenarios are either functionally redundant or that more significant disruption of canonical and/or non-canonical spliceosome pathways is not tolerated by the cells.

Although spliceosome mutations are frequent in MDS, the consequences of these aberrations have not been fully elucidated yet. Among the immediate questions raised by these findings are, why different spliceosome mutations result in varied morphological changes as observed in MDS patients. It is noteworthy that spliceosome mutations are largely restricted to proteins involved in the recognition of 3′ splice-site. Therefore, the disruption of 3′ spice-site recognition as a result of these mutations could potentially alter the splicing in a systematic manner by either affecting exon recognition/utilization or perhaps activating alternative splice junctions. As a result, dysfunctional spliceosome could generate downstream novel protein isoforms with altered function, stimulate ectopic expression of tissue-inappropriate protein isoforms or drive premature activation of nonsense-mediated decay of proteins resulting in quantitative changes in gene expression which can causes downstream changes in the cells. In support of this, a strong genotype-phenotype correlation exists between *SF3B1* mutations and the presence of ringed sideroblasts. This highly specific relationship has been confirmed in all subsequent studies (Papaemmanuil et al., 2013). This makes *SF3B1* the first gene to be strongly associated with a specific morphologic feature in myeloid malignancies, and implies that these aberrations are not benign spectators and are central to the pathogenesis of the disease. Recent studies have shown that *SRSF2* and *U2AF1* mutations alter the normal sequence-

specific RNA binding specificity of these gene, thereby modifying the recognition of specific exonic splicing enhancer motifs to drive recurrent mis-splicing of key genes involved in the haematopoiesis (Ilagan et al., 2015; Kim et al., 2015). A study investigating the relationship of *SF3B1* and ring sideroblasts reported the presence of ring sideroblasts in *SF3B1*$^{+/-}$ mice (Visconte et al., 2012b). A long term follow-up of the SF3B1$^{+/-}$ mice also revealed macrocytic anaemia, extramedullary haematopoiesis, and thrombocytosis (Visconte et al., 2014). *In-vitro* targeting of the *SF3B1* using meayamycin, a pharmacologic inhibitor of the splicing factor 3b complex (Albert et al., 2009), induced the formation of ring sideroblasts in healthy bone marrow cells (Visconte et al., 2012b). Although, the percentage of ring sideroblasts in MDS patients is highly correlated to the *SF3B1* mutant allele burden, however, the amount of ring sideroblasts still ranges considerably. Therefore, the above variability suggests that additional factors must play a role in the ring sideroblasts phenotypic expression. Initial transcriptome analysis including ours, however, have not revealed compelling, systematic evidence suggesting patterns of mutation-associated splicing abnormalities, therefore raising the possibility that these *SF3B1* mutations may exert their function by disrupting alternative non-splicing pathways. In addition to their role in precursor mRNA processing, *SF3B1* has been shown to influence histone modifications via physical interactions with polycomb group proteins, influence cellular response to DNA damage and activate transcription via direct and indirect regulation of RNA polymerase elongation (Isono et al., 2005; Kfir et al., 2015; Te Raa et al., 2015). In addition, *SF3B1* contains several cyclin-dependent kinase (cdk) consensus phosphorylation sites within its N terminus. In *in-vitro* studies, it has been shown that cyclin E-cdk2 targets *SF3B1* and phosphorylation of *SF3B1* occurs prior or during the catalytic step II of the splicing pathway (Seghezzi et al., 1998) and subsequently gets dephosphorylated by PP1/PP2A phosphatases concomitant with the second step of splicing (Wahl et al., 2009). Interestingly, in MDS patients, a link between cell cycle progression and splicing has also been seen where particular isoforms of the CDC25C cell cycle regulator have been shown in association with disease (Caudill et al., 2008). Weather the cause of the unusual pattern of iron deposition within the erythroid cells is solely as a result of defects in the splicing machinery or through an alternative *SF3B1* target gene(s) or pathway(s) remains to be determined.

In our study and all other studies published to date, up to 15% of the RARS patients have no evidence of mutations in *SF3B1*. This could be partly due to inaccuracies in the diagnosis of these patients related to difficulties in the identification of ring sideroblasts. However, the possibility exists that other unidentified spliceosome genes may also be mutated in these patients which may lead to formation of ring sideroblasts.

Somatic mutations are known to occur in the hematopoietic stem cells (HSCs) at a lower frequency during normal DNA replication. Although most of the mutations are rapidly corrected by DNA repair mechanisms or perhaps kept under control by immune system, those aberrations that persist are propagated during HSC self-renewal. Each individual stem cell accumulates exonic mutations at a rate of approximately 0.13 per year, therefore by the age of 60, each HSC would have accumulated approximately upto 8 mutations with the potential to have an effect on the protein function. MDS being an old age-related disease has long been considered to be an oligoclonal stem cell disorder. However, the experimental evidence describing the existence of such human cancer stem cells (or MDS initiating stem cells) remain extensively contested and *in-vivo* tracing of candidate MDS initiating stem cells (MDS-ISC) in such patients has not been successful. Here, we demonstrated for the first time that *SF3B1* mutation(s) arise from early $CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$ HSCs providing the strongest evidence for the existence of MDS-ISCs in RARS that leads to the expansion and dominance of a heterozygous clone, propagating to its myeloid progeny. This is the first report to isolate the $CD49f^{+}$ HSC fraction from MDS bone marrow and study the genetics as well as engraftment kinetics of these cells. The existence of such MDS-ISCs has also recently been revealed in 5q-MDS patients (Woll et al., 2014). Notably, *SF3B1* mutations were not detected in lymphoid compartment, unlike chronic lymphocytic leukaemia (CLL) where mutations are solely observed in lymphoid cells. In contrast to RARS/RCMD-RS, *SF3B1* mutation in CLL occurs during the clonal progression of the disease and determine the emergence of more aggressive subclones. The available evidence suggests that *SF3B1* mutations are very unlikely to be founding mutations and must be originating in the late progenitors in CLL. Nonetheless, the notion that lymphoid and myeloid diseases can share mutant genes is potentially interesting, as shown by a study on *TET2* mutations in malignant lymphomas (Quivoron et al., 2011). This is at least consistent with the concept that the

same effected genes can cause 2 completely different diseases depending upon the cell type in which it occurs.

The significance of functional follow-up in xenograft models is recognized but still expensive as well as time consuming. It is also worth to note that very few of the novel mutations receive functional validation in *in-vitro* and *in-vivo* studies due to the fact that unlike leukaemia, engraftment and propagation of MDS-derived cells has proven difficult (Nilsson et al., 2002; Thanopoulou et al., 2004). Although acquired sideroblastic anaemia has been recognized as a relatively homogenous disease from the early days of experimental cancer research, the intra-tumor genetic heterogeneity has not been studied in detail. For the first time, genetically distinguishable sub-clonal populations in a mixture of cancer stem cells were revealed by single-nucleus and single-cell sequencing (Hou et al., 2012; Navin et al., 2011; Xu et al., 2012). Using combination of techniques including single cell clonogenic assays, we have clearly demonstrated the co-existence of genetically divergent sub-clones within *SF3B1* mutant cell population. This intra-tumor heterogeneity has long been suggested as a common feature in both haematological malignancies such as AML and solid tumours. Recent studies have evaluated the clonal relationships among primary and metastatic cancers (Campbell et al., 2010; Gerlinger et al., 2012; Navin et al., 2011; Yachida et al., 2010). These studies demonstrated that seeding metastases need few, if any, additional driver lesions beyond those found in the primary malignant cells. The basic premise of this long-standing idea is that a cell that is endowed with an advantageous mutation produces a progeny that has a survival advantage over other cells that lack this mutation. Our data suggest that *SF3B1* is such a 'first', or "gatekeeping" mutation which enables the outgrowth of malignant clone over the cohabitating normal HSCs initially as a microscopic clone. Consequently, the progeny of this malignant cell with increased fitness does flourish and produces a clonal population that dominates. Subsequently, this mutant clone also starts to have downstream effects on the clonal erythroid cells, ultimately causing the ring sideroblasts, through a direct or indirect mechanism. This is also evident from our *in-vivo* (NSG mice) as well as *in-vitro* experiments, where *SF3B1* mutant present in a mix population of cells (*SF3B1* mutant/wildtype), eventually grows and outcompetes normal HSCs. Over time, additional 'co-operating/advantageous' mutations arise (in epigenetic regulator or

transcription/cell signalling genes), endowing a further growth advantage to a cell within the clone. As unique sub-clones arise within the *SF3B1* clone, different outcomes are possible: less fit sub-clones can be completely lost with the most fit sub-clone dominating, or many minor sub-clones can persist alongside the dominant clone, forming reservoirs from which evolution (or disease progression) can continue as evident from our NSG xenograft experiments which shows more potent clones such as “SF3B1- DNMT3A” can out-compete “SF3B1- JAK2” clones. This was also seen in our patient data where continuous acquisition of mutations and clonal expansion during the progression of the disease caused an assemblage of subclones that were represented by a branching multi-clonal as well as ancestral evolutionary tree. It has been estimated that each acquired “driver” mutation gives only a small selective growth advantage, on the order of a 0.4% increase in the difference between cell birth and cell death. Over the years, however, this slight increase, compounded once or twice per week, can result in a large expansion of the malignant clones (Bozic et al., 2010). With this model in mind, as disease progresses, clonal hierarchies become shallower within genetic subclones. In some cases, once the disease has progressed to such an advanced state, the frequency of ‘lethal’ malignant clones may be so high that the tumour subclone essentially becomes functionally homogeneous without evidence of a hierarchy as evidenced in some patients (MDS2) in our study. However, the finding that all subclones were detected in the founding clone suggests that *SF3B1* is a “gatekeeping-driver” mutation which requires additional ‘potent’ aberrations for disease progression/transformation. Overall, such models where sub-clones persist and/or contribute to independent phylogenetic lineage trees within the founding clone are reminiscent of the branching evolution that Darwin proposed as leading to increased fitness and overall robustness of a particular organism.

Another major factor to consider here is the role of the bone marrow non-malignant (or malignant) cell elements associated with tumour cells, referred to collectively as the bone marrow microenvironment (or niches) (Morrison and Scadden, 2014). The juxtaposition of a malignant cell carrying the *SF3B1* mutation within the micro-environment might have an influence on the function as well as the evolution of that clone, resulting in significant variation in path it takes during the course of the disease. The complexity imposed by the microenvironment is amplified due to crosstalk

between these malignant cells and the bone marrow microenvironment. This bone marrow microenvironment pressure might also be the defining feature which determines if the *SF3B1* clone follows a branching multi-clonal or ancestral evolutionary path during the course of the disease.

From a conceptual standpoint, it is clear that therapy failure and MDS disease recurrence (or transformation to AML) are not just simply due to the acquisition of new mutations in a normal stem cell. Rather, the surviving malignant clones must escape the therapy, remain dormant and also have regenerative potential in order to cause the disease relapse/transformation; therefore, cells contributing to recurrence must behave like MDS-ISCs or disease transforming cells. Certainly, the only critical cells in a malignant bone marrow are the ones that are responsible for long-term clonal maintenance and expansion; any other cell will ultimately lead to clonal exhaustion (Greaves, 2013). Our data suggests that a cell like HSC with highest self-renewal capacity and carrying an *SF3B1* mutation seems to survive various treatments during the course of the disease, remains dormant and then acquires additional mutations (such as *FLT3* or *RUNX1*) enabling the disease to transform at the HSC level. However, this is not true in all the cases as evident from our sequencing analysis of the sequential samples in MDS-AML patient (MDS2) which uncovered the clonal evolution in hematopoietic progeny (with the acquisition of *SUV420H1* mutation) and selection of this malignant driving clone leading to leukemic transformation. This paradigm illustrates the concept of a switch from “HSC disease” to a “Progenitor disease”. This is consistent with the thinking in the field that pre-leukemia must initiate in a self-renewing HSC but leukemia may be generated from mutations acquired in a more mature progenitor cell in some patients. Therefore, our data suggests that in some cases minor HSC populations carrying the *SF3B1* mutations are escaping the treatment (later acquires additional mutations) while as in other cases the proliferation pressure caused as a result of the treatment results in the acquisition of additional mutation at the progenitor level turning them into leukaemia-initiating cells. In support of this, there is a strong evidence emerging to suggest a link between stemness and therapy resistance in MDS with del(5q), multiple myeloma, lymphoid malignancies, breast cancer and colon cancer where studies show that cancer initiating stem cell carrying the mutations are more resistant to therapy (Diehn et al., 2009; Leung-

Hagesteijn et al., 2013; Tehranchi et al., 2010; Zhang et al., 2010). Together, these studies highlight the interplay between cancer stem cell properties and genetics that drive clinical parameters such as response to therapy and ultimately survival of the malignant clone which could also be true for *SF3B1* mutant MDS patients.

The emerging evidence linking stemness to prognosis and therapy failure in other haematological malignancies, indicates that therapeutic targeting of *SF3B1* clone along with the elements of stemness might be an effective means to eradicate MDS-ISCs and prevent disease recurrence or transformation. Although, it still unclear as to how stemness is regulated, several factors of the self-renewal machinery such as BMI1 have been strongly linked to self-renewal and have been implicated in the maintenance of stem cells in several tissues (Lessard and Sauvageau, 2003; Molofsky et al., 2003). A recent study by *Kreso et al*., demonstrated that by targeting BMI1, inhibits the ability of colorectal stem cells to self-renew, resulting in abrogation of their tumorigenic potential in mouse xenografts (Kreso et al., 2014). These studies point to the need for the development of combination therapies (using nanoparticle technology) to attempt targeting of *SF3B1* mutations along with other predicted components of the self-renewal machinery as well as altering the microenvironment supporting these cells. It is important to note that because stemness-associated factors are likely to be shared between normal stem cells and MDS-ISCs, successful eradication of MDS-ISCs will require understanding to what extent MDS-ISCs differ from normal stem cells to minimize the impact of therapies on these HSCs. The application of single cell RNA-Seq might be able to delineate this difference in the transcription and signalling networks (hence identifying the cell surface markers/signals) that must exist between normal stem cells and *SF3B1* mutant MDS-ISCs.

Furthermore, delineating genetic from nongenetic stemness influences is also important factor which can help to tease apart the aspects of *SF3B1* mutant MDS disease. Moreover, the dynamics of interactions between sub-clones, each with their own superimposed developmental hierarchy - weather they compete or are co-depend upon each other and hence coordinate clonal evolution, needs to be elucidated. Darwinian Theory dictates how the evolution creates biodiversity and this in-turn makes an entire ecosystem robust. Therefore, clonal diversity within the hierarchy at

the genetic and functional level together with their ‘cross-talk’ with the microenvironment could also increase clonal fitness, allowing these *SF3B1* malignant cells to offset survival pressures imposed by therapy. Therefore, more effective therapy strategies will require gaining insight into this diversity.

Future Work

In recent years the heterogeneity of MDS in-general has been deciphered further by the discovery of defects in spliceosome genes as reported here and elsewhere. Our knowledge has taken an enormous step forward with the unearthing of potential connections between RNA splicing defects, the specific structural features of pre-RNAs and many of the processes such as RNA processing kinetics, genome stability as well as cell cycle processes.

This study shows for the first time shows the presence of *SF3B1* mutants MDS-ICs in MDS patients with ring sideroblasts and how these cells inherent a clonal superiority over the cohabitating cells in the bone marrow. The main question which arises is how could basal splicing machinery itself possibly be deranged without causing catastrophic cellular consequences? The mRNA splicing process itself is so fundamental to every normal cells' behaviour that any serious defect in the small nuclear ribonucleoproteins (snRNPs) or the >200 associated proteins that catalyses splicing should be lethal. Therefore, are there other non-spliceosome pathways through which SF3B1 is acting and causing the phenotypic changes i.e. ring sideroblasts in MDS patients. Therefore, there are two major lines of investigation that could be undertaken

1. Transcriptome analysis of *SF3B1* mutant bone marrow cells at a single cell level

2. Genome editing the *SF3B1* mutations using CRISPR.

References

- Abdel-Wahab, O., Adli, M., LaFave, L.M., Gao, J., Hricik, T., Shih, A.H., Pandey, S., Patel, J.P., Chung, Y.R., Koche, R.*, et al.* (2012). ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. Cancer cell *22*, 180-193.
- Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R.*, et al.* (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144-147.
- Abu Kar, S., Jankowska, A.M., Makishima, H., Visconte, V., Jerez, A., Sugimoto, Y., Muramatsu, H., Traina, F., Afable, M., Guinta, K.*, et al.* (2012). Spliceosomal gene mutations are frequent events in the diverse mutational spectrum of chronic myelomonocytic leukemia but largely absent in juvenile myelomonocytic leukemia. Haematologica.
- Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.
- Albert, B.J., McPherson, P.A., O'Brien, K., Czaicki, N.L., Destefino, V., Osman, S., Li, M., Day, B.W., Grabowski, P.J., Moore, M.J.*, et al.* (2009). Meayamycin inhibits pre-messenger RNA splicing and exhibits picomolar activity against multidrug-resistant cells. Molecular cancer therapeutics *8*, 2308-2318.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X.*, et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503-511.
- Anand, S., Stedham, F., Beer, P., Gudgin, E., Ortmann, C.A., Bench, A., Erber, W., Green, A.R., and Huntly, B.J. (2011). Effects of the JAK2 mutation on the hematopoietic stem and progenitor compartment in human myeloproliferative neoplasms. Blood *118*, 177-181.
- Anderson, C.P., Shen, M., Eisenstein, R.S., and Leibold, E.A. (2012). Mammalian iron metabolism and its control by iron regulatory proteins. Biochimica et biophysica acta *1823*, 1468-1483.
- Anderson, K., Lutz, C., van Delft, F.W., Bateman, C.M., Guo, Y., Colman, S.M., Kempski, H., Moorman, A.V., Titley, I., Swansbury, J.*, et al.* (2011). Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature *469*, 356-361.
- Arosio, P., Ingrassia, R., and Cavadini, P. (2009). Ferritins: a family of molecules for iron storage, antioxidation and more. Biochimica et biophysica acta *1790*, 589-599.
- Asou, H., Matsui, H., Ozaki, Y., Nagamachi, A., Nakamura, M., Aki, D., and Inaba, T. (2009). Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. Biochemical and biophysical research communications *383*, 245-251.
- Bains, A., Luthra, R., Medeiros, L.J., and Zuo, Z. (2011). FLT3 and NPM1 mutations in myelodysplastic syndromes: Frequency and potential value for predicting progression to acute myeloid leukemia. Am J Clin Pathol *135*, 62-69.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.
- Barlow, J.L., Drynan, L.F., Hewett, D.R., Holmes, L.R., Lorenzo-Abalde, S., Lane, A.L., Jolin, H.E., Pannell, R., Middleton, A.J., Wong, S.H.*, et al.* (2010). A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. Nature medicine *16*, 59-66.
- Barrett, A.J., and Sloand, E. (2009). Autoimmune mechanisms in the pathophysiology of myelodysplastic syndromes and their clinical relevance. Haematologica *94*, 449-451.
- Bassaganyas, L., Bea, S., Escaramis, G., Tornador, C., Salaverria, I., Zapata, L., Drechsel, O., Ferreira, P.G., Rodriguez-Santiago, B., Tubio, J.M.*, et al.* (2013).

Sporadic and reversible chromothripsis in chronic lymphocytic leukemia revealed by longitudinal genomic analysis. Leukemia *27*, 2376-2379.

- Baum, C.M., Weissman, I.L., Tsukamoto, A.S., Buckle, A.M., and Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. Proceedings of the National Academy of Sciences of the United States of America *89*, 2804-2808.
- Becker, A.J., Mc, C.E., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452-454.
- Bejar, R., Levine, R., and Ebert, B.L. (2011). Unraveling the molecular pathophysiology of myelodysplastic syndromes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *29*, 504-515.
- Bejar, R., Stevenson, K.E., Caughey, B.A., Abdel-Wahab, O., Steensma, D.P., Galili, N., Raza, A., Kantarjian, H., Levine, R.L., Neuberg, D.*, et al.* (2012). Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *30*, 3376-3382.
- Benetti, R., Gonzalo, S., Jaco, I., Schotta, G., Klatt, P., Jenuwein, T., and Blasco, M.A. (2007). Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. The Journal of cell biology *178*, 925-936.
- Bergmann, A.K., Campagna, D.R., McLoughlin, E.M., Agarwal, S., Fleming, M.D., Bottomley, S.S., and Neufeld, E.J. (2010). Systematic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. Pediatr Blood Cancer *54*, 273-278.
- Bhatia, M., Wang, J.C., Kapp, U., Bonnet, D., and Dick, J.E. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. Proceedings of the National Academy of Sciences of the United States of America *94*, 5320-5325.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L.*, et al.* (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research *42*, W252-258.
- Boissinot, M., Garand, R., Hamidou, M., and Hermouet, S. (2006). The JAK2-V617F mutation and essential thrombocythemia features in a subset of patients with refractory anemia with ring sideroblasts (RARS). Blood *108*, 1781-1782.
- Bonanno, G., Mariotti, A., Procoli, A., Corallo, M., Scambia, G., Pierelli, L., and Rutella, S. (2009). Interleukin-21 induces the differentiation of human umbilical cord blood CD34-lineage- cells into pseudomature lytic NK cells. BMC immunology *10*, 46.
- Bonnal, S., Martinez, C., Forch, P., Bachi, A., Wilm, M., and Valcarcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. Molecular cell *32*, 81-95.
- Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature medicine *3*, 730-737.
- Bosma, G.C., Custer, R.P., and Bosma, M.J. (1983). A severe combined immunodeficiency mutation in the mouse. Nature *301*, 527-530.
- Bosma, M.J., and Carroll, A.M. (1991). The SCID mouse mutant: definition, characterization, and potential uses. Annual review of immunology *9*, 323-350.
- Boultwood, J., Pellagatti, A., Nikpour, M., Pushkaran, B., Fidler, C., Cattan, H., Littlewood, T.J., Malcovati, L., Della Porta, M.G., Jadersten, M.*, et al.* (2008). The role of the iron transporter ABCB7 in refractory anemia with ring sideroblasts. PloS one *3*, e1970.
- Boultwood, J., Perry, J., Pellagatti, A., Fernandez-Mercado, M., Fernandez-Santamaria, C., Calasanz, M.J., Larrayoz, M.J., Garcia-Delgado, M., Giagounidis, A.,

Malcovati, L.*, et al.* (2010). Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. Leukemia *24*, 1062-1065.

- Boultwood, J., and Wainscoat, J.S. (2001). Clonality in the myelodysplastic syndromes. Int J Hematol *73*, 411-415.
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. Proceedings of the National Academy of Sciences of the United States of America *107*, 18545-18550.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes & development *20*, 1123-1136.
- Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. (2003). EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. The EMBO journal *22*, 5323-5335.
- Breuer, W., Ronson, A., Slotki, I.N., Abramov, A., Hershko, C., and Cabantchik, Z.I. (2000). The assessment of serum nontransferrin-bound iron in chelation therapy and iron supplementation. Blood *95*, 2975-2982.
- Busque, L., Patel, J.P., Figueroa, M.E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A.*, et al.* (2012). Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nature genetics *44*, 1179-1181.
- Camaschella, C. (2009). Hereditary sideroblastic anemias: pathophysiology, diagnosis, and treatment. Semin Hematol *46*, 371-377.
- Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L.*, et al.* (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature *467*, 1109-1113.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.
- Caputi, M., and Zahler, A.M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. The Journal of biological chemistry *276*, 43850-43859.
- Cashman, J.D., Lapidot, T., Wang, J.C., Doedens, M., Shultz, L.D., Lansdorp, P., Dick, J.E., and Eaves, C.J. (1997). Kinetic evidence of the regeneration of multilineage hematopoiesis from primitive cells in normal human bone marrow transplanted into immunodeficient mice. Blood *89*, 4307-4316.
- Cass, D.M., and Berglund, J.A. (2006). The SF3b155 N-terminal domain is a scaffold important for splicing. Biochemistry *45*, 10092-10101.
- Caudill, J.S., Porcher, J.C., and Steensma, D.P. (2008). Aberrant pre-mRNA splicing of a highly conserved cell cycle regulator, CDC25C, in myelodysplastic syndromes. Leukemia & lymphoma *49*, 989-993.
- Cazzola, M., May, A., Bergamaschi, G., Cerani, P., Rosti, V., and Bishop, D.F. (2000). Familial-skewed X-chromosome inactivation as a predisposing factor for late-onset X-linked sideroblastic anemia in carrier females. Blood *96*, 4363-4365.
- Cazzola, M., and Invernizzi, R. (2011). Ring sideroblasts and sideroblastic anemias. Haematologica *96*, 789-792.
- Cazzola, M., Rossi, M., and Malcovati, L. (2012). Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood.
- Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y.*, et al.* (2012). Dnmt3a is essential for hematopoietic stem cell differentiation. Nature genetics *44*, 23-31.

- Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M.*, et al.* (2011). Initial genome sequencing and analysis of multiple myeloma. Nature *471*, 467-472.
- Chen, C.D., Kobayashi, R., and Helfman, D.M. (1999). Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. Genes & development *13*, 593-606.
- Chen, G., Zeng, W., Miyazato, A., Billings, E., Maciejewski, J.P., Kajigaya, S., Sloand, E.M., and Young, N.S. (2004). Distinctive gene expression profiles of CD34 cells from patients with myelodysplastic syndrome characterized by specific chromosomal abnormalities. Blood *104*, 4210-4218.
- Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature reviews Molecular cell biology *10*, 741-754.
- Cho, Y.S., Kim, E.J., Park, U.H., Sin, H.S., and Um, S.J. (2006). Additional sex comb-like 1 (ASXL1), in cooperation with SRC-1, acts as a ligand-dependent coactivator for retinoic acid receptor. The Journal of biological chemistry *281*, 17588-17598.
- Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. Molecular and cellular biology *19*, 69-77.
- Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K., Li, X.S., Woon, E.C., Yang, M.*, et al.* (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO reports *12*, 463-469.
- Christiansen, D.H., Andersen, M.K., and Pedersen-Bjergaard, J. (2001). Mutations with loss of heterozygosity of p53 are common in therapy-related myelodysplasia and acute myeloid leukemia after exposure to alkylating agents and significantly associated with deletion or loss of 5q, a complex karyotype, and a poor prognosis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *19*, 1405-1413.
- Christiansen, D.H., Andersen, M.K., and Pedersen-Bjergaard, J. (2004). Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. Blood *104*, 1474-1481.
- Cingolani, G., Petosa, C., Weis, K., and Muller, C.W. (1999). Structure of importin-beta bound to the IBB domain of importin-alpha. Nature *399*, 221-229.
- Civin, C.I., Strauss, L.C., Brovall, C., Fackler, M.J., Schwartz, J.F., and Shaper, J.H. (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. Journal of immunology *133*, 157-165.
- Claessens, Y.E., Park, S., Dubart-Kupperschmitt, A., Mariot, V., Garrido, C., Chretien, S., Dreyfus, F., Lacombe, C., Mayeux, P., and Fontenay, M. (2005). Rescue of early-stage myelodysplastic syndrome-deriving erythroid precursors by the ectopic expression of a dominant-negative form of FADD. Blood *105*, 4035-4042.
- Cotter, P.D., May, A., Li, L., Al-Sabah, A.I., Fitzsimons, E.J., Cazzola, M., and Bishop, D.F. (1999). Four new mutations in the erythroid-specific 5-aminolevulinate synthase (ALAS2) gene causing X-linked sideroblastic anemia: increased pyridoxine responsiveness after removal of iron overload by phlebotomy and coinheritance of hereditary hemochromatosis. Blood *93*, 1757-1769.
- Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della-Valle, V., Couronne, L., Scourzic, L., Chesnais, V.*, et al.* (2012a). Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood.
- Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della Valle, V., Couronne, L., Scourzic, L., Chesnais, V.*, et al.* (2012b).

Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood *119*, 3211-3218.

- Damm, F., Thol, F., Kosmider, O., Kade, S., Loffeld, P., Dreyfus, F., Stamatoullas-Bastard, A., Tanguy-Schmidt, A., Beyne-Rauzy, O., de Botton, S.*, et al.* (2011). SF3B1 mutations in myelodysplastic syndromes: clinical associations and prognostic implications. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK.
- Damm, F., Thol, F., Kosmider, O., Kade, S., Loffeld, P., Dreyfus, F., Stamatoullas-Bastard, A., Tanguy-Schmidt, A., Beyne-Rauzy, O., de Botton, S.*, et al.* (2012c). SF3B1 mutations in myelodysplastic syndromes: clinical associations and prognostic implications. Leukemia *26*, 1137-1140.
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C.*, et al.* (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739-744.
- Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell *150*, 12-27.
- de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Molecular cell *12*, 525-532.
- Del Rey, M., Benito, R., Fontanillo, C., Campos-Laborie, F.J., Janusz, K., Velasco-Hernandez, T., Abaigar, M., Hernandez, M., Cuello, R., Borrego, D.*, et al.* (2015). Deregulation of Genes Related to Iron and Mitochondrial Metabolism in Refractory Anemia with Ring Sideroblasts. PloS one *10*, e0126555.
- Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A.*, et al.* (2009). Mutation in TET2 in myeloid cancers. The New England journal of medicine *360*, 2289-2301.
- Diehn, M., Cho, R.W., Lobo, N.A., Kalisky, T., Dorie, M.J., Kulp, A.N., Qian, D., Lam, J.S., Ailles, L.E., Wong, M.*, et al.* (2009). Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature *458*, 780-783.
- Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D.*, et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature *481*, 506-510.
- Dohner, K., Brown, J., Hehmann, U., Hetzel, C., Stewart, J., Lowther, G., Scholl, C., Frohling, S., Cuneo, A., Tsui, L.C.*, et al.* (1998). Molecular cytogenetic characterization of a critical region in bands 7q35-q36 commonly deleted in malignant myeloid disorders. Blood *92*, 4031-4035.
- Douglas, A.S., and Dacie, J.V. (1953). The incidence and significance of iron-containing granules in human erythrocytes and their precursors. Journal of clinical pathology *6*, 307-313.
- Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nature immunology *11*, 585-593.
- Dowdy, C.R., Xie, R., Frederick, D., Hussain, S., Zaidi, S.K., Vradii, D., Javed, A., Li, X., Jones, S.N., Lian, J.B.*, et al.* (2010). Definitive hematopoiesis requires Runx1 C-terminal-mediated subnuclear targeting and transactivation. Human molecular genetics *19*, 1048-1057.
- Dunbar, A.J., Gondek, L.P., O'Keefe, C.L., Makishima, H., Rataul, M.S., Szpurka, H., Sekeres, M.A., Wang, X.F., McDevitt, M.A., and Maciejewski, J.P. (2008). 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. Cancer research *68*, 10349-10357.

- Ebert, B.L., Pretz, J., Bosco, J., Chang, C.Y., Tamayo, P., Galili, N., Raza, A., Root, D.E., Attar, E., Ellis, S.R.*, et al.* (2008). Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. Nature *451*, 335-339.
- Endler, G., Greinix, H., Winkler, K., Mitterbauer, G., and Mannhalter, C. (1999). Genetic fingerprinting in mouthwashes of patients after allogeneic bone marrow transplantation. Bone marrow transplantation *24*, 95-98.
- Eppert, B.L., Wortham, B.W., Flury, J.L., and Borchers, M.T. (2013). Functional characterization of T cell populations in a mouse model of chronic obstructive pulmonary disease. Journal of immunology *190*, 1331-1340.
- Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A.*, et al.* (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nature genetics *42*, 722-726.
- Esteller, M. (2008). Epigenetics in cancer. The New England journal of medicine *358*, 1148-1159.
- Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.
- Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. Genes & development *17*, 419-437.
- Feuring-Buske, M., Gerhard, B., Cashman, J., Humphries, R.K., Eaves, C.J., and Hogge, D.E. (2003). Improved engraftment of human acute myeloid leukemia progenitor cells in beta 2-microglobulin-deficient NOD/SCID mice and in NOD/SCID mice transgenic for human growth factors. Leukemia *17*, 760-763.
- Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F.*, et al.* (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer cell *18*, 553-567.
- Figueroa, M.E., Skrabanek, L., Li, Y., Jiemjit, A., Fandy, T.E., Paietta, E., Fernandez, H., Tallman, M.S., Greally, J.M., Carraway, H.*, et al.* (2009). MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. Blood *114*, 3448-3458.
- Fisher, C.L., Pineault, N., Brookes, C., Helgason, C.D., Ohta, H., Bodner, C., Hess, J.L., Humphries, R.K., and Brock, H.W. (2010). Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. Blood *115*, 38-46.
- Fisher, C.L., Randazzo, F., Humphries, R.K., and Brock, H.W. (2006). Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. Gene *369*, 109-118.
- Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K.*, et al.* (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature genetics *37*, 391-400.
- Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.
- Galm, O., Herman, J.G., and Baylin, S.B. (2006). The fundamental role of epigenetics in hematopoietic malignancies. Blood Rev *20*, 1-13.
- Galy, A., Travis, M., Cen, D., and Chen, B. (1995). Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. Immunity *3*, 459-473.
- Gelsi-Boyer, V., Trouplin, V., Adelaide, J., Bonansea, J., Cervera, N., Carbuccia, N., Lagarde, A., Prebet, T., Nezri, M., Sainty, D.*, et al.* (2009). Mutations of polycomb-

associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. British journal of haematology *145*, 788-800.

- Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P.*, et al.* (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine *366*, 883-892.
- Giarratana, M.C., Kobari, L., Lapillonne, H., Chalmers, D., Kiger, L., Cynober, T., Marden, M.C., Wajcman, H., and Douay, L. (2005). Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. Nature biotechnology *23*, 69-74.
- Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R.*, et al.* (2011). Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer cell *19*, 138-152.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. Science *300*, 980-984.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2005). Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. Molecular cell *17*, 869-883.
- Goldstein, E.R., Ziegenfuss, T., Kalman, D., Kreider, R., Campbell, B., Wilborn, C., Taylor, L., Willoughby, D., Stout, J., Graves, B.S.*, et al.* (2010). International society of sports nutrition position stand: caffeine and performance. J Int Soc Sports Nutr *7*, 5.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A.*, et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531-537.
- Gondek, L.P., Tiu, R., O'Keefe, C.L., Sekeres, M.A., Theil, K.S., and Maciejewski, J.P. (2008). Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. Blood *111*, 1534-1542.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Molecular and cellular biology *18*, 4752-4760.
- Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E.*, et al.* (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nature genetics *44*, 53-57.
- Greaves, M. (2013). Cancer stem cells as 'units of selection'. Evolutionary applications *6*, 102-108.
- Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306-313.
- Greco, M., Capello, D., Bruscaggin, A., Spina, V., Rasi, S., Monti, S., Ciardullo, C., Cresta, S., Fangazio, M., Gaidano, G.*, et al.* (2012). Analysis of SF3B1 mutations in monoclonal B-cell lymphocytosis. Hematological oncology.
- Greenberg, P., Cox, C., LeBeau, M.M., Fenaux, P., Morel, P., Sanz, G., Sanz, M., Vallespi, T., Hamblin, T., Oscier, D.*, et al.* (1997). International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood *89*, 2079-2088.
- Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Sole, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F.*, et al.* (2012). Revised international prognostic scoring system for myelodysplastic syndromes. Blood *120*, 2454-2465.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C.*, et al.* (2007). Patterns of somatic mutation in human cancer genomes. Nature *446*, 153-158.
- Grisendi, S., Bernardi, R., Rossi, M., Cheng, K., Khandker, L., Manova, K., and Pandolfi, P.P. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. Nature *437*, 147-153.

- Gross, S., Cairns, R.A., Minden, M.D., Driggers, E.M., Bittinger, M.A., Jang, H.G., Sasaki, M., Jin, S., Schenkein, D.P., Su, S.M.*, et al.* (2010). Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. The Journal of experimental medicine *207*, 339-344.
- Groves, M.R., and Barford, D. (1999). Topological characteristics of helical repeat proteins. Current opinion in structural biology *9*, 383-389.
- Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R.*, et al.* (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494-504.
- Haase, D., Germing, U., Schanz, J., Pfeilstocker, M., Nosslinger, T., Hildebrandt, B., Kundgen, A., Lubbert, M., Kunzmann, R., Giagounidis, A.A.*, et al.* (2007). New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. Blood *110*, 4385-4395.
- Haferlach, C., Bacher, U., Haferlach, T., Dicker, F., Alpermann, T., Kern, W., and Schnittger, S. (2011). The inv(3)(q21q26)/t(3;3)(q21;q26) is frequently accompanied by alterations of the RUNX1, KRAS and NRAS and NF1 genes and mediates adverse prognosis both in MDS and in AML: a study in 39 cases of MDS or AML. Leukemia *25*, 874-877.
- Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T.*, et al.* (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241-247.
- Hahn, C.N., and Scott, H.S. (2012). Spliceosome mutations in hematopoietic malignancies. Nature genetics *44*, 9-10.
- Hao, Q.L., Zhu, J., Price, M.A., Payne, K.J., Barsky, L.W., and Crooks, G.M. (2001). Identification of a novel, human multilymphoid progenitor in cord blood. Blood *97*, 3683-3690.
- Harada, H., Harada, Y., Niimi, H., Kyo, T., Kimura, A., and Inaba, T. (2004). High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia. Blood *103*, 2316-2324.
- Harada, Y., and Harada, H. (2009). Molecular pathways mediating MDS/AML with focus on AML1/RUNX1 point mutations. Journal of cellular physiology *220*, 16-20.
- Heinrichs, S., Kulkarni, R.V., Bueso-Ramos, C.E., Levine, R.L., Loh, M.L., Li, C., Neuberg, D., Kornblau, S.M., Issa, J.P., Gilliland, D.G.*, et al.* (2009). Accurate detection of uniparental disomy and microdeletions by SNP array analysis in myelodysplastic syndromes with normal cytogenetics. Leukemia *23*, 1605-1613.
- Hellstrom-Lindberg, E., and Cazzola, M. (2008). The role of JAK2 mutations in RARS and other MDS. Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program, 52-59.
- Herrera-Merchan, A., Arranz, L., Ligos, J.M., de Molina, A., Dominguez, O., and Gonzalez, S. (2012). Ectopic expression of the histone methyltransferase Ezh2 in haematopoietic stem cells causes myeloproliferative disease. Nature communications *3*, 623.
- Hesselton, R.M., Greiner, D.L., Mordes, J.P., Rajan, T.V., Sullivan, J.L., and Shultz, L.D. (1995). High levels of human peripheral blood mononuclear cell engraftment and enhanced susceptibility to human immunodeficiency virus type 1 infection in NOD/LtSz-scid/scid mice. J Infect Dis *172*, 974-982.
- Hoebeke, I., De Smedt, M., Stolz, F., Pike-Overzet, K., Staal, F.J., Plum, J., and Leclercq, G. (2007). T-, B- and NK-lymphoid, but not myeloid cells arise from human CD34(+)CD38(-)CD7(+) common lymphoid progenitors expressing lymphoid-specific genes. Leukemia *21*, 311-319.

- Hogan, C.J., Shpall, E.J., and Keller, G. (2002). Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice. Proceedings of the National Academy of Sciences of the United States of America *99*, 413-418.
- Horiike, S., Kita-Sasai, Y., Nakao, M., and Taniwaki, M. (2003). Configuration of the TP53 gene as an independent prognostic parameter of myelodysplastic syndrome. Leukemia & lymphoma *44*, 915-922.
- Horton, S.J., and Huntly, B.J. (2012). Recent advances in acute myeloid leukemia stem cell biology. Haematologica *97*, 966-974.
- Hou, Y., Song, L., Zhu, P., Zhang, B., Tao, Y., Xu, X., Li, F., Wu, K., Liang, J., Shao, D*., et al.* (2012). Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell *148*, 873-885.
- Howe, R.B., Porwit-MacDonald, A., Wanat, R., Tehranchi, R., and Hellstrom-Lindberg, E. (2004). The WHO classification of MDS does make a difference. Blood *103*, 3265-3270.
- Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome research *25*, 14-26.
- Ingram, W., Lea, N.C., Cervera, J., Germing, U., Fenaux, P., Cassinat, B., Kiladjian, J.J., Varkonyi, J., Antunovic, P., Westwood, N.B*., et al.* (2006). The JAK2 V617F mutation identifies a subgroup of MDS patients with isolated deletion 5q and a proliferative bone marrow. Leukemia *20*, 1319-1321.
- Ishikawa, F., Yasukawa, M., Lyons, B., Yoshida, S., Miyamoto, T., Yoshimoto, G., Watanabe, T., Akashi, K., Shultz, L.D., and Harada, M. (2005). Development of functional human blood and immune systems in NOD/SCID/IL2 receptor {gamma} chain(null) mice. Blood *106*, 1565-1573.
- Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N*., et al.* (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nature biotechnology *25*, 1315-1321.
- Isono, K., Abe, K., Tomaru, Y., Okazaki, Y., Hayashizaki, Y., and Koseki, H. (2001). Molecular cloning, genetic mapping, and expression of the mouse Sf3b1 (SAP155) gene for the U2 snRNP component of spliceosome. Mamm Genome *12*, 192-198.
- Isono, K., Mizutani-Koseki, Y., Komori, T., Schmidt-Zachmann, M.S., and Koseki, H. (2005). Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev *19*, 536-541.
- Issa, J.P. (2010). Epigenetic changes in the myelodysplastic syndrome. Hematology/oncology clinics of North America *24*, 317-330.
- Ito, M., Hiramatsu, H., Kobayashi, K., Suzue, K., Kawahata, M., Hioki, K., Ueyama, Y., Koyanagi, Y., Sugamura, K., Tsuji, K*., et al.* (2002). NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells. Blood *100*, 3175-3182.
- Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.J*., et al.* (2012). Detectable clonal mosaicism and its relationship to aging and cancer. Nature genetics *44*, 651-658.
- Jadersten, M., Malcovati, L., Dybedal, I., Della Porta, M.G., Invernizzi, R., Montgomery, S.M., Pascutto, C., Porwit, A., Cazzola, M., and Hellstrom-Lindberg, E. (2008). Erythropoietin and granulocyte-colony stimulating factor treatment associated with improved survival in myelodysplastic syndrome. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 3607-3613.
- Jadersten, M., Saft, L., Smith, A., Kulasekararaj, A., Pomplun, S., Gohring, G., Hedlund, A., Hast, R., Schlegelberger, B., Porwit, A*., et al.* (2011). TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. Journal

of clinical oncology : official journal of the American Society of Clinical Oncology *29*, 1971-1979.

- Janssen, J.W., Buschle, M., Layton, M., Drexler, H.G., Lyons, J., van den Berghe, H., Heimpel, H., Kubanek, B., Kleihauer, E., Mufti, G.J.*, et al.* (1989). Clonal analysis of myelodysplastic syndromes: evidence of multipotent stem cell origin. Blood *73*, 248-254.
- Jerez, A., Gondek, L.P., Jankowska, A.M., Makishima, H., Przychodzen, B., Tiu, R.V., O'Keefe, C.L., Mohamedali, A.M., Batista, D., Sekeres, M.A.*, et al.* (2012). Topography, clinical, and genomic correlates of 5q myeloid malignancies revisited. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *30*, 1343-1349.
- Jiang, Y., Dunbar, A., Gondek, L.P., Mohan, S., Rataul, M., O'Keefe, C., Sekeres, M., Saunthararajah, Y., and Maciejewski, J.P. (2009). Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. Blood *113*, 1315-1325.
- Jin, G., Reitman, Z.J., Spasojevic, I., Batinic-Haberle, I., Yang, J., Schmidt-Kittler, O., Bigner, D.D., and Yan, H. (2011). 2-hydroxyglutarate production, but not dominant negative function, is conferred by glioma-derived NADP-dependent isocitrate dehydrogenase mutations. PloS one *6*, e16812.
- Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A.*, et al.* (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science *321*, 1801-1806.
- Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H.*, et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nature chemical biology *3*, 576-583.
- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. Nature reviews Genetics *12*, 715-729.
- Kamminga, L.M., Bystrykh, L.V., de Boer, A., Houwer, S., Douma, J., Weersing, E., Dontje, B., and de Haan, G. (2006). The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood *107*, 2170-2179.
- Kaneko, K., Furuyama, K., Fujiwara, T., Kobayashi, R., Ishida, H., Harigae, H., and Shibahara, S. (2013). Identification of the novel erythroid-specific enhancer for ALAS2 gene and its loss-of-function mutation associated with congenital sideroblastic anemia. Haematologica.
- Kang, Y., Chao, N.J., and Aversa, F. (2008). Unmanipulated or CD34 selected haplotype mismatched transplants. Current opinion in hematology *15*, 561-567.
- Kawamoto, H., Ikawa, T., Masuda, K., Wada, H., and Katsura, Y. (2010). A map for lineage restriction of progenitors during hematopoiesis: the essence of the myeloid-based model. Immunological reviews *238*, 23-36.
- Kfir, N., Lev-Maor, G., Glaich, O., Alajem, A., Datta, A., Sze, S.K., Meshorer, E., and Ast, G. (2015). SF3B1 Association with Chromatin Determines Splicing Outcomes. Cell reports *11*, 618-629.
- Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E.*, et al.* (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer cell *27*, 617-630.
- Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proceedings of the National Academy of Sciences of the United States of America *108*, 13564-13569.
- Kita-Sasai, Y., Horiike, S., Misawa, S., Kaneko, H., Kobayashi, M., Nakao, M., Nakagawa, H., Fujii, H., and Taniwaki, M. (2001). International prognostic scoring

system and TP53 mutations are independent prognostic indicators for patients with myelodysplastic syndrome. British journal of haematology *115*, 309-312.

- Klco, J.M., Spencer, D.H., Miller, C.A., Griffith, M., Lamprecht, T.L., O'Laughlin, M., Fronick, C., Magrini, V., Demeter, R.T., Fulton, R.S.*, et al.* (2014). Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer cell *25*, 379-392.
- Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25*, 2283-2285.
- Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research *22*, 568-576.
- Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661-672.
- Konig, H., Matter, N., Bader, R., Thiele, W., and Muller, F. (2007). Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. Cell *131*, 718-729.
- Korbel, J.O., and Campbell, P.J. (2013). Criteria for inference of chromothripsis in cancer genomes. Cell *152*, 1226-1236.
- Kosmider, O., Gelsi-Boyer, V., Slama, L., Dreyfus, F., Beyne-Rauzy, O., Quesnel, B., Hunault-Berger, M., Slama, B., Vey, N., Lacombe, C.*, et al.* (2010). Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms. Leukemia *24*, 1094-1096.
- Koury, M.J., and Bondurant, M.C. (1990). Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells. Science *248*, 378-381.
- Kreso, A., van Galen, P., Pedley, N.M., Lima-Fernandes, E., Frelin, C., Davis, T., Cao, L., Baiazitov, R., Du, W., Sydorenko, N.*, et al.* (2014). Self-renewal as a therapeutic target in human colorectal cancer. Nature medicine *20*, 29-36.
- Krummel, D.A., Nagai, K., and Oubridge, C. (2010). Structure of spliceosomal ribonucleoproteins. F1000 Biol Rep *2*.
- Kuwasako, K., Dohmae, N., Inoue, M., Shirouzu, M., Taguchi, S., Guntert, P., Seraphin, B., Muto, Y., and Yokoyama, S. (2008). Complex assembly mechanism and an RNA-binding mode of the human p14-SF3b155 spliceosomal protein complex identified by NMR solution structure and functional analyses. Proteins *71*, 1617-1636.
- Lane, S.W., Sykes, S.M., Al-Shahrour, F., Shterental, S., Paktinat, M., Lo Celso, C., Jesneck, J.L., Ebert, B.L., Williams, D.A., and Gilliland, D.G. (2010). The Apc(min) mouse has altered hematopoietic stem cell function and provides a model for MPD/MDS. Blood *115*, 3489-3497.
- Langemeijer, S.M., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A.*, et al.* (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nature genetics *41*, 838-842.
- Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C.*, et al.* (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nature genetics *44*, 642-650.
- Le Beau, M.M., Espinosa, R., 3rd, Davis, E.M., Eisenbart, J.D., Larson, R.A., and Green, E.D. (1996). Cytogenetic and molecular delineation of a region of chromosome 7 commonly deleted in malignant myeloid diseases. Blood *88*, 1930-1935.
- Lee, S.W., Cho, Y.S., Na, J.M., Park, U.H., Kang, M., Kim, E.J., and Um, S.J. (2010a). ASXL1 represses retinoic acid receptor-mediated transcription through associating with HP1 and LSD1. The Journal of biological chemistry *285*, 18-29.

- Lee, W., Jiang, Z., Liu, J., Haverty, P.M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K.P., Bhatt, D.*, et al.* (2010b). The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature *465*, 473-477.
- Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M.*, et al.* (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155-165.
- Lessard, J., and Sauvageau, G. (2003). Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature *423*, 255-260.
- Leung-Hagesteijn, C., Erdmann, N., Cheung, G., Keats, J.J., Stewart, A.K., Reece, D.E., Chung, K.C., and Tiedemann, R.E. (2013). Xbp1s-negative tumor B cells and pre-plasmablasts mediate therapeutic proteasome inhibitor resistance in multiple myeloma. Cancer cell *24*, 289-304.
- Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J.*, et al.* (2010). DNMT3A mutations in acute myeloid leukemia. The New England journal of medicine *363*, 2424-2433.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
- Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509-4518.
- Liang, H., Fairman, J., Claxton, D.F., Nowell, P.C., Green, E.D., and Nagarajan, L. (1998). Molecular anatomy of chromosome 7q deletions in myeloid neoplasms: evidence for multiple critical loci. Proceedings of the National Academy of Sciences of the United States of America *95*, 3781-3785.
- Lill, R., and Muhlenhoff, U. (2008). Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. Annual review of biochemistry *77*, 669-700.
- Liu, R., Wang, X., Chen, G.Y., Dalerba, P., Gurney, A., Hoey, T., Sherlock, G., Lewicki, J., Shedden, K., and Clarke, M.F. (2007). The prognostic role of a gene signature from tumorigenic breast-cancer cells. The New England journal of medicine *356*, 217-226.
- Lowry, P.A., Shultz, L.D., Greiner, D.L., Hesselton, R.M., Kittler, E.L., Tiarks, C.Y., Rao, S.S., Reilly, J., Leif, J.H., Ramshaw, H.*, et al.* (1996). Improved engraftment of human cord blood stem cells in NOD/LtSz-scid/scid mice after irradiation or multiple-day injections into unirradiated recipients. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation *2*, 15-23.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.
- Maguire, A., Hellier, K., Hammans, S., and May, A. (2001). X-linked cerebellar ataxia and sideroblastic anaemia associated with a missense mutation in the ABC7 gene predicting V411L. British journal of haematology *115*, 910-917.
- Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. Cell stem cell *1*, 635-645.
- Majewski, I.J., Ritchie, M.E., Phipson, B., Corbin, J., Pakusch, M., Ebert, A., Busslinger, M., Koseki, H., Hu, Y., Smyth, G.K.*, et al.* (2010). Opposing roles of polycomb repressive complexes in hematopoietic stem and progenitor cells. Blood *116*, 731-739.
- Makishima, H., Jankowska, A.M., Tiu, R.V., Szpurka, H., Sugimoto, Y., Hu, Z., Sauntharajah, Y., Guinta, K., Keddache, M.A., Putnam, P.*, et al.* (2010a). Novel

homo- and hemizygous mutations in EZH2 in myeloid malignancies. Leukemia *24*, 1799-1804.

- Makishima, H., Rataul, M., Gondek, L.P., Huh, J., Cook, J.R., Theil, K.S., Sekeres, M.A., Kuczkowski, E., O'Keefe, C., and Maciejewski, J.P. (2010b). FISH and SNP-A karyotyping in myelodysplastic syndromes: improving cytogenetic detection of del(5q), monosomy 7, del(7q), trisomy 8 and del(20q). Leukemia research *34*, 447-453.
- Makishima, H., Visconte, V., Sakaguchi, H., Jankowska, A.M., Abu Kar, S., Jerez, A., Przychodzen, B., Bupathi, M., Guinta, K., Afable, M.G.*, et al.* (2012). Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood.
- Malcovati, L., Della Porta, M.G., Pietra, D., Boveri, E., Pellagatti, A., Galli, A., Travaglino, E., Brisci, A., Rumi, E., Passamonti, F.*, et al.* (2009). Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis. Blood *114*, 3538-3545.
- Malcovati, L., Papaemmanuil, E., Bowen, D.T., Boultwood, J., Della Porta, M.G., Pascutto, C., Travaglino, E., Groves, M.J., Godfrey, A.L., Ambaglio, I.*, et al.* (2011). Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood *118*, 6239-6246.
- Manz, M.G., Miyamoto, T., Akashi, K., and Weissman, I.L. (2002). Prospective isolation of human clonogenic common myeloid progenitors. Proceedings of the National Academy of Sciences of the United States of America *99*, 11872-11877.
- Mardis, E.R., Ding, L., Dooling, D.J., Larson, D.E., McLellan, M.D., Chen, K., Koboldt, D.C., Fulton, R.S., Delehaunty, K.D., McGrath, S.D.*, et al.* (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med *361*, 1058-1066.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.
- Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nature reviews Molecular cell biology *15*, 108-121.
- Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P., and Tavtigian, S.V. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic acids research *34*, 1317-1325.
- Matheny, C.J., Speck, M.E., Cushing, P.R., Zhou, Y., Corpora, T., Regan, M., Newman, M., Roudaia, L., Speck, C.L., Gu, T.L.*, et al.* (2007). Disease mutations in RUNX1 and RUNX2 create nonfunctional, dominant-negative, or hypomorphic alleles. The EMBO journal *26*, 1163-1175.
- Mazurier, F., Doedens, M., Gan, O.I., and Dick, J.E. (2003). Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. Nature medicine *9*, 959-963.
- McCullough, A.J., and Berget, S.M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Molecular and cellular biology *17*, 4562-4571.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research *20*, 1297-1303.
- Medyouf, H., Mossner, M., Jann, J.C., Nolte, F., Raffel, S., Herrmann, C., Lier, A., Eisen, C., Nowak, V., Zens, B.*, et al.* (2014). Myelodysplastic cells in patients reprogram mesenchymal stromal cells to establish a transplantable stem cell niche disease unit. Cell stem cell *14*, 824-837.
- Metzeler, K.H., Maharry, K., Radmacher, M.D., Mrozek, K., Margeson, D., Becker, H., Curfman, J., Holland, K.B., Schwind, S., Whitman, S.P.*, et al.* (2011). TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia:

a Cancer and Leukemia Group B study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *29*, 1373-1381.

- Meyerson, M., Gabriel, S., and Getz, G. (2010). Advances in understanding cancer genomes through second-generation sequencing. Nature reviews Genetics *11*, 685-696.
- Mian, S.A., Smith, A.E., Kulasekararaj, A.G., Kizilors, A., Mohamedali, A.M., Lea, N.C., Mitsopoulos, K., Ford, K., Nasser, E., Seidl, T.*, et al.* (2013). Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica *98*, 1058-1066.
- Michallet, M., Philip, T., Philip, I., Godinot, H., Sebban, C., Salles, G., Thiebaut, A., Biron, P., Lopez, F., Mazars, P.*, et al.* (2000). Transplantation with selected autologous peripheral blood CD34+Thy1+ hematopoietic stem cells (HSCs) in multiple myeloma: impact of HSC dose on engraftment, safety, and immune reconstitution. Experimental hematology *28*, 858-870.
- Mohamedali, A., Gaken, J., Twine, N.A., Ingram, W., Westwood, N., Lea, N.C., Hayden, J., Donaldson, N., Aul, C., Gattermann, N.*, et al.* (2007). Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes. Blood *110*, 3365-3373.
- Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.K., Clarke, M.F., and Morrison, S.J. (2003). Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature *425*, 962-967.
- Mongkonsritragoon, W., Letendre, L., and Li, C.Y. (1998). Multiple lymphoid nodules in bone marrow have the same clonality as underlying myelodysplastic syndrome recognized with fluorescent in situ hybridization technique. Am J Hematol *59*, 252-257.
- Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer cell *20*, 11-24.
- Mordes, D., Luo, X., Kar, A., Kuo, D., Xu, L., Fushimi, K., Yu, G., Sternberg, P., Jr., and Wu, J.Y. (2006). Pre-mRNA splicing and retinitis pigmentosa. Molecular vision *12*, 1259-1271.
- Morrison, S.J., and Scadden, D.T. (2014). The bone marrow niche for haematopoietic stem cells. Nature *505*, 327-334.
- Mosier, D.E., Stell, K.L., Gulizia, R.J., Torbett, B.E., and Gilmore, G.L. (1993). Homozygous scid/scid;beige/beige mice have low levels of spontaneous or neonatal T cell-induced B cell generation. The Journal of experimental medicine *177*, 191-194.
- Mufti, G.J., Bennett, J.M., Goasguen, J., Bain, B.J., Baumann, I., Brunning, R., Cazzola, M., Fenaux, P., Germing, U., Hellstrom-Lindberg, E.*, et al.* (2008). Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. Haematologica *93*, 1712-1717.
- Murray, L., Chen, B., Galy, A., Chen, S., Tushinski, R., Uchida, N., Negrin, R., Tricot, G., Jagannath, S., Vesole, D.*, et al.* (1995). Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood. Blood *85*, 368-378.
- Napier, I., Ponka, P., and Richardson, D.R. (2005). Iron trafficking in the mitochondrion: novel pathways revealed by disease. Blood *105*, 1867-1874.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D.*, et al.* (2011). Tumour evolution inferred by single-cell sequencing. Nature *472*, 90-94.
- Negrin, R.S., Atkinson, K., Leemhuis, T., Hanania, E., Juttner, C., Tierney, K., Hu, W.W., Johnston, L.J., Shizurn, J.A., Stockerl-Goldstein, K.E.*, et al.* (2000).

Transplantation of highly purified CD34+Thy-1+ hematopoietic stem cells in patients with metastatic breast cancer. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation *6*, 262-271.

- Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H., and Boehm, T. (1994). New member of the winged-helix protein family disrupted in mouse and rat nude mutations. Nature *372*, 103-107.
- Netz, D.J., Pierik, A.J., Stumpfig, M., Muhlenhoff, U., and Lill, R. (2007). The Cfd1-Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol. Nature chemical biology *3*, 278-286.
- Neuenkirchen, N., Chari, A., and Fischer, U. (2008). Deciphering the assembly pathway of Sm-class U snRNPs. FEBS letters *582*, 1997-2003.
- Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C.*, et al.* (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics *42*, 790-793.
- Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tonnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T.*, et al.* (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nature genetics *42*, 665-667.
- Nikpour, M., Pellagatti, A., Liu, A., Karimi, M., Malcovati, L., Gogvadze, V., Forsblom, A.M., Wainscoat, J.S., Cazzola, M., Zhivotovsky, B.*, et al.* (2010). Gene expression profiling of erythroblasts from refractory anaemia with ring sideroblasts (RARS) and effects of G-CSF. British journal of haematology *149*, 844-854.
- Nikpour, M., Scharenberg, C., Liu, A., Conte, S., Karimi, M., Mortera-Blanco, T., Giai, V., Fernandez-Mercado, M., Papaemmanuil, E., Hogstrand, K.*, et al.* (2012). The transporter ABCB7 is a mediator of the phenotype of acquired refractory anemia with ring sideroblasts. Leukemia.
- Nilsson, L., Astrand-Grundstrom, I., Anderson, K., Arvidsson, I., Hokland, P., Bryder, D., Kjeldsen, L., Johansson, B., Hellstrom-Lindberg, E., Hast, R.*, et al.* (2002). Involvement and functional impairment of the CD34(+)CD38(-)Thy-1(+) hematopoietic stem cell pool in myelodysplastic syndromes with trisomy 8. Blood *100*, 259-267.
- Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science *333*, 218-221.
- Nottrott, S., Urlaub, H., and Luhrmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. The EMBO journal *21*, 5527-5538.
- Novotna, B., Neuwirtova, R., Siskova, M., and Bagryantseva, Y. (2008). DNA instability in low-risk myelodysplastic syndromes: refractory anemia with or without ring sideroblasts. Human molecular genetics *17*, 2144-2149.
- Ohgami, R.S., Campagna, D.R., Greer, E.L., Antiochos, B., McDonald, A., Chen, J., Sharp, J.J., Fujiwara, Y., Barker, J.E., and Fleming, M.D. (2005). Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells. Nature genetics *37*, 1264-1269.
- Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644-649.
- Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell *84*, 321-330.
- Oltean, S., and Bates, D.O. (2014). Hallmarks of alternative splicing in cancer. Oncogene *33*, 5311-5318.

- Orr, S.J., Boutz, D.R., Wang, R., Chronis, C., Lea, N.C., Thayaparan, T., Hamilton, E., Milewicz, H., Blanc, E., Mufti, G.J.*, et al.* (2012). Proteomic and protein interaction network analysis of human T lymphocytes during cell-cycle entry. Mol Syst Biol *8*, 573.
- Otto, E.A., Hurd, T.W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., Patil, S.B., Levy, S., Ghosh, A.K., Murga-Zamalloa, C.A.*, et al.* (2010). Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. Nature genetics *42*, 840-850.
- Padgett, R.A. (2012). New connections between splicing and human disease. Trends Genet *28*, 147-154.
- Pang, W.W., Pluvinage, J.V., Price, E.A., Sridhar, K., Arber, D.A., Greenberg, P.L., Schrier, S.L., Park, C.Y., and Weissman, I.L. (2013). Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. Proceedings of the National Academy of Sciences of the United States of America *110*, 3011-3016.
- Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C.*, et al.* (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med *365*, 1384-1395.
- Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A.*, et al.* (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699.
- Parikh, C., Subrahmanyam, R., and Ren, R. (2006). Oncogenic NRAS rapidly and efficiently induces CMML- and AML-like diseases in mice. Blood *108*, 2349-2357.
- Park, S., Grabar, S., Kelaidi, C., Beyne-Rauzy, O., Picard, F., Bardet, V., Coiteux, V., Leroux, G., Lepelley, P., Daniel, M.T.*, et al.* (2008). Predictive factors of response and survival in myelodysplastic syndrome treated with erythropoietin and G-CSF: the GFM experience. Blood *111*, 574-582.
- Park, U.H., Yoon, S.K., Park, T., Kim, E.J., and Um, S.J. (2011). Additional sex comb-like (ASXL) proteins 1 and 2 play opposite roles in adipogenesis via reciprocal regulation of peroxisome proliferator-activated receptor {gamma}. The Journal of biological chemistry *286*, 1354-1363.
- Paschka, P., Schlenk, R.F., Gaidzik, V.I., Habdank, M., Kronke, J., Bullinger, L., Spath, D., Kayser, S., Zucknick, M., Gotze, K.*, et al.* (2010). IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *28*, 3636-3643.
- Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO journal *23*, 4061-4071.
- Patnaik, M.M., Lasho, T.L., Hodnefield, J.M., Knudson, R.A., Ketterling, R.P., Garcia-Manero, G., Steensma, D.P., Pardanani, A., Hanson, C.A., and Tefferi, A. (2012). SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. Blood *119*, 569-572.
- Pedersen-Bjergaard, J., Christiansen, D.H., Desta, F., and Andersen, M.K. (2006). Alternative genetic pathways and cooperating genetic abnormalities in the pathogenesis of therapy-related myelodysplasia and acute myeloid leukemia. Leukemia *20*, 1943-1949.
- Pellagatti, A., Cazzola, M., Giagounidis, A.A., Malcovati, L., Porta, M.G., Killick, S., Campbell, L.J., Wang, L., Langford, C.F., Fidler, C.*, et al.* (2006). Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood *108*, 337-345.

- Ponka, P., Beaumont, C., and Richardson, D.R. (1998). Function and regulation of transferrin and ferritin. Seminars in hematology *35*, 35-54.
- Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research *35*, D61-65.
- Puente, X.S., Pinyol, M., Quesada, V., Conde, L., Ordonez, G.R., Villamor, N., Escaramis, G., Jares, P., Bea, S., Gonzalez-Diaz, M.*, et al.* (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature *475*, 101-105.
- Query, C.C., Strobel, S.A., and Sharp, P.A. (1996). Three recognition events at the branch-site adenine. The EMBO journal *15*, 1392-1402.
- Quintana, E., Shackleton, M., Sabel, M.S., Fullen, D.R., Johnson, T.M., and Morrison, S.J. (2008). Efficient tumour formation by single human melanoma cells. Nature *456*, 593-598.
- Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H.*, et al.* (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer cell *20*, 25-38.
- Rathinam, C., Thien, C.B., Langdon, W.Y., Gu, H., and Flavell, R.A. (2008). The E3 ubiquitin ligase c-Cbl restricts development and functions of hematopoietic stem cells. Genes & development *22*, 992-997.
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333-i339.
- Raza, A., and Galili, N. (2012). The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. Nature reviews Cancer *12*, 849-859.
- Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-111.
- Richardson, D.R., Lane, D.J., Becker, E.M., Huang, M.L., Whitnall, M., Suryo Rahmanto, Y., Sheftel, A.D., and Ponka, P. (2010). Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. Proceedings of the National Academy of Sciences of the United States of America *107*, 10775-10782.
- Ritchie, D.B., Schellenberg, M.J., and MacMillan, A.M. (2009). Spliceosome structure: piece by piece. Biochimica et biophysica acta *1789*, 624-633.
- Robertson, K.D. (2005). DNA methylation and human disease. Nature reviews Genetics *6*, 597-610.
- Robson-Dixon, N.D., and Garcia-Blanco, M.A. (2004). MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. The Journal of biological chemistry *279*, 29075-29084.
- Sanada, M., Suzuki, T., Shih, L.Y., Otsu, M., Kato, M., Yamazaki, S., Tamura, A., Honda, H., Sakata-Yanagimoto, M., Kumano, K.*, et al.* (2009). Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature *460*, 904-908.
- Sargin, B., Choudhary, C., Crosetto, N., Schmidt, M.H., Grundler, R., Rensinghoff, M., Thiessen, C., Tickenbrock, L., Schwable, J., Brandts, C.*, et al.* (2007). Flt3-dependent transformation by inactivating c-Cbl mutations in AML. Blood *110*, 1004-1012.
- Saur, S.J., Sangkhae, V., Geddis, A.E., Kaushansky, K., and Hitchcock, I.S. (2010). Ubiquitination and degradation of the thrombopoietin receptor c-Mpl. Blood *115*, 1254-1263.
- Schneider, R.K., Adema, V., Heckl, D., Jaras, M., Mallo, M., Lord, A.M., Chu, L.P., McConkey, M.E., Kramann, R., Mullally, A.*, et al.* (2014). Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. Cancer cell *26*, 509-520.

- Schmidt-Mende, J., Tehranchi, R., Forsblom, A.M., Joseph, B., Christensson, B., Fadeel, B., Zhivotovsky, B., and Hellstrom-Lindberg, E. (2001). Granulocyte colony-stimulating factor inhibits Fas-triggered apoptosis in bone marrow cells isolated from patients with refractory anemia with ringed sideroblasts. Leukemia *15*, 742-751.
- Scott, B.L., and Deeg, H.J. (2010). Myelodysplastic syndromes. Annual review of medicine *61*, 345-358.
- Seghezzi, W., Chua, K., Shanahan, F., Gozani, O., Reed, R., and Lees, E. (1998). Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. Molecular and cellular biology *18*, 4526-4536.
- Shackleton, M., Quintana, E., Fearon, E.R., and Morrison, S.J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell *138*, 822-829.
- Shah, S.P., Morin, R.D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J.*, et al.* (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature *461*, 809-813.
- Shaw, G.C., Cope, J.J., Li, L., Corson, K., Hersey, C., Ackermann, G.E., Gwynn, B., Lambert, A.J., Wingert, R.A., Traver, D.*, et al.* (2006). Mitoferrin is essential for erythroid iron assimilation. Nature *440*, 96-100.
- Sheftel, A.D., Richardson, D.R., Prchal, J., and Ponka, P. (2009). Mitochondrial iron metabolism and sideroblastic anemia. Acta haematologica *122*, 120-133.
- Shen, L., Kantarjian, H., Guo, Y., Lin, E., Shan, J., Huang, X., Berry, D., Ahmed, S., Zhu, W., Pierce, S.*, et al.* (2010). DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *28*, 605-613.
- Shih, L.Y., Huang, C.F., Wang, P.N., Wu, J.H., Lin, T.L., Dunn, P., and Kuo, M.C. (2004). Acquisition of FLT3 or N-ras mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK *18*, 466-475.
- Shultz, L.D., Ishikawa, F., and Greiner, D.L. (2007). Humanized mice in translational biomedical research. Nat Rev Immunol *7*, 118-130.
- Shultz, L.D., Lang, P.A., Christianson, S.W., Gott, B., Lyons, B., Umeda, S., Leiter, E., Hesselton, R., Wagar, E.J., Leif, J.H.*, et al.* (2000). NOD/LtSz-Rag1null mice: an immunodeficient and radioresistant model for engraftment of human hematolymphoid cells, HIV infection, and adoptive transfer of NOD mouse diabetogenic T cells. Journal of immunology *164*, 2496-2507.
- Shultz, L.D., Lyons, B.L., Burzenski, L.M., Gott, B., Chen, X., Chaleff, S., Kotb, M., Gillies, S.D., King, M., Mangada, J.*, et al.* (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. Journal of immunology *174*, 6477-6489.
- Shultz, L.D., Schweitzer, P.A., Christianson, S.W., Gott, B., Schweitzer, I.B., Tennent, B., McKenna, S., Mobraaten, L., Rajan, T.V., Greiner, D.L.*, et al.* (1995). Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. Journal of immunology *154*, 180-191.
- Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nature reviews Molecular cell biology *10*, 697-708.
- Smallwood, A., Esteve, P.O., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes & development *21*, 1169-1178.
- Smith, A.E., Mian, S.A., Kulasekararaj, A.G., Mohamedali, A.M., and Mufti, G.M. (2011). Whole Exome Sequencing Reveals Acquired SF3B1 Mutations Defining Patients with Acquired Idiopathic Sideroblastic Anaemia Blood (ASH Annual Meeting) 2011 *633*.

- Smith, A.E., Mohamedali, A.M., Kulasekararaj, A., Lim, Z., Gaken, J., Lea, N.C., Przychodzen, B., Mian, S.A., Nasser, E.E., Shooter, C.*, et al.* (2010). Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low abundance mutant clones with early origins, but indicates no definite prognostic value. Blood.
- Smith, E., and Shilatifard, A. (2007). The A, B, Gs of silencing. Genes & development *21*, 1141-1144.
- Smith, O.P., Hann, I.M., Woodward, C.E., and Brockington, M. (1995). Pearson's marrow/pancreas syndrome: haematological features associated with deletion and duplication of mitochondrial DNA. British journal of haematology *90*, 469-472.
- Sole, F., Luno, E., Sanzo, C., Espinet, B., Sanz, G.F., Cervera, J., Calasanz, M.J., Cigudosa, J.C., Milla, F., Ribera, J.M.*, et al.* (2005). Identification of novel cytogenetic markers with prognostic significance in a series of 968 patients with primary myelodysplastic syndromes. Haematologica *90*, 1168-1178.
- Starczynowski, D.T., Kuchenbauer, F., Argiropoulos, B., Sung, S., Morin, R., Muranyi, A., Hirst, M., Hogge, D., Marra, M., Wells, R.A.*, et al.* (2010). Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. Nature medicine *16*, 49-58.
- Stark, H., and Luhrmann, R. (2006). Cryo-electron microscopy of spliceosomal components. Annual review of biophysics and biomolecular structure *35*, 435-457.
- Steensma, D.P., and Tefferi, A. (2008). JAK2 V617F and ringed sideroblasts: not necessarily RARS-T. Blood *111*, 1748.
- Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A.*, et al.* (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27-40.
- Sun, H., and Chasin, L.A. (2000). Multiple splicing defects in an intronic false exon. Molecular and cellular biology *20*, 6414-6425.
- Swanton, C. (2012). Intratumor heterogeneity: evolution through space and time. Cancer research *72*, 4875-4882.
- Szpurka, H., Tiu, R., Murugesan, G., Aboudola, S., Hsi, E.D., Theil, K.S., Sekeres, M.A., and Maciejewski, J.P. (2006). Refractory anemia with ringed sideroblasts associated with marked thrombocytosis (RARS-T), another myeloproliferative condition characterized by JAK2 V617F mutation. Blood *108*, 2173-2181.
- Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development *19*, 815-826.
- Tanaka, S., Miyagi, S., Sashida, G., Chiba, T., Yuan, J., Mochizuki-Kashio, M., Suzuki, Y., Sugano, S., Nakaseko, C., Yokote, K.*, et al.* (2012). Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia. Blood *120*, 1107-1117.
- Taussig, D.C., Miraki-Moud, F., Anjos-Afonso, F., Pearce, D.J., Allen, K., Ridler, C., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S.G.*, et al.* (2008). Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. Blood *112*, 568-575.
- Taussig, D.C., Vargaftig, J., Miraki-Moud, F., Griessinger, E., Sharrock, K., Luke, T., Lillington, D., Oakervee, H., Cavenagh, J., Agrawal, S.G.*, et al.* (2010). Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. Blood *115*, 1976-1984.
- Tavtigian, S.V., Deffenbaugh, A.M., Yin, L., Judkins, T., Scholl, T., Samollow, P.B., de Silva, D., Zharkikh, A., and Thomas, A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. Journal of medical genetics *43*, 295-305.

- Te Raa, G.D., Derks, I.A., Navrkalova, V., Skowronska, A., Moerland, P.D., van Laar, J., Oldreive, C., Monsuur, H., Trbusek, M., Malcikova, J.*, et al.* (2015). The impact of SF3B1 mutations in CLL on the DNA-damage response. Leukemia *29*, 1133-1142.
- Tefferi, A., and Vardiman, J.W. (2009). Myelodysplastic syndromes. N Engl J Med *361*, 1872-1885.
- Tehranchi, R., Fadeel, B., Forsblom, A.M., Christensson, B., Samuelsson, J., Zhivotovsky, B., and Hellstrom-Lindberg, E. (2003). Granulocyte colony-stimulating factor inhibits spontaneous cytochrome c release and mitochondria-dependent apoptosis of myelodysplastic syndrome hematopoietic progenitors. Blood *101*, 1080-1086.
- Tehranchi, R., Invernizzi, R., Grandien, A., Zhivotovsky, B., Fadeel, B., Forsblom, A.M., Travaglino, E., Samuelsson, J., Hast, R., Nilsson, L.*, et al.* (2005). Aberrant mitochondrial iron distribution and maturation arrest characterize early erythroid precursors in low-risk myelodysplastic syndromes. Blood *106*, 247-253.
- Tehranchi, R., Woll, P.S., Anderson, K., Buza-Vidas, N., Mizukami, T., Mead, A.J., Astrand-Grundstrom, I., Strombeck, B., Horvat, A., Ferry, H.*, et al.* (2010). Persistent malignant stem cells in del(5q) myelodysplasia in remission. The New England journal of medicine *363*, 1025-1037.
- Thanopoulou, E., Cashman, J., Kakagianne, T., Eaves, A., Zoumbos, N., and Eaves, C. (2004). Engraftment of NOD/SCID-beta2 microglobulin null mice with multilineage neoplastic cells from patients with myelodysplastic syndrome. Blood *103*, 4285-4293.
- Thol, F., Kade, S., Schlarmann, C., Loffeld, P., Morgan, M., Krauter, J., Wlodarski, M.W., Kolking, B., Wichmann, M., Gorlich, K.*, et al.* (2012a). Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood *119*, 3578-3584.
- Thol, F., Kade, S., Schlarmann, C., Loffeld, P., Morgan, M., Krauter, J., Wlodarski, M.W., Kolking, B., Wichmann, M., Gorlich, K.*, et al.* (2012b). Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood.
- Thol, F., Weissinger, E.M., Krauter, J., Wagner, K., Damm, F., Wichmann, M., Gohring, G., Schumann, C., Bug, G., Ottmann, O.*, et al.* (2010). IDH1 mutations in patients with myelodysplastic syndromes are associated with an unfavorable prognosis. Haematologica *95*, 1668-1674.
- Till, J.E., and Mc, C.E. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation research *14*, 213-222.
- Tiu, R.V., Gondek, L.P., O'Keefe, C.L., Huh, J., Sekeres, M.A., Elson, P., McDevitt, M.A., Wang, X.F., Levis, M.J., Karp, J.E.*, et al.* (2009). New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *27*, 5219-5226.
- Tosi, S., Scherer, S.W., Giudici, G., Czepulkowski, B., Biondi, A., and Kearney, L. (1999). Delineation of multiple deleted regions in 7q in myeloid disorders. Genes, chromosomes & cancer *25*, 384-392.
- Traggiai, E., Chicha, L., Mazzucchelli, L., Bronz, L., Piffaretti, J.C., Lanzavecchia, A., and Manz, M.G. (2004). Development of a human adaptive immune system in cord blood cell-transplanted mice. Science *304*, 104-107.
- Tryndyak, V.P., Kovalchuk, O., and Pogribny, I.P. (2006). Loss of DNA methylation and histone H4 lysine 20 trimethylation in human breast cancer cells is associated with aberrant expression of DNA methyltransferase 1, Suv4-20h2 histone methyltransferase and methyl-binding proteins. Cancer biology & therapy *5*, 65-70.
- van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T.*, et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

- van der Feltz, C., Anthony, K., Brilot, A., and Pomeranz Krummel, D.A. (2012). Architecture of the spliceosome. Biochemistry *51*, 3321-3333.
- Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.
- Visconte, V., Makishima, H., Jankowska, A., Szpurka, H., Traina, F., Jerez, A., O'Keefe, C., Rogers, H.J., Sekeres, M.A., Maciejewski, J.P.*, et al.* (2011). SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. Leukemia.
- Visconte, V., Makishima, H., Jankowska, A., Szpurka, H., Traina, F., Jerez, A., O'Keefe, C., Rogers, H.J., Sekeres, M.A., Maciejewski, J.P.*, et al.* (2012a). SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK *26*, 542-545.
- Visconte, V., Rogers, H.J., Singh, J., Barnard, J., Bupathi, M., Traina, F., McMahon, J., Makishima, H., Szpurka, H., Jankowska, A.*, et al.* (2012b). SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. Blood *120*, 3173-3186.
- Visconte, V., Tabarroki, A., Zhang, L., Parker, Y., Hasrouni, E., Mahfouz, R., Isono, K., Koseki, H., Sekeres, M.A., Sauntharajah, Y.*, et al.* (2014). Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk Myelodysplastic syndromes with ring sideroblasts. Journal of hematology & oncology *7*, 89.
- Vogel, W., Scheding, S., Kanz, L., and Brugger, W. (2000). Clinical applications of CD34(+) peripheral blood progenitor cells (PBPC). Stem cells *18*, 87-92.
- Vose, J.M., Bierman, P.J., Lynch, J.C., Atkinson, K., Juttner, C., Hanania, C.E., Bociek, G., and Armitage, J.O. (2001). Transplantation of highly purified CD34+Thy-1+ hematopoietic stem cells in patients with recurrent indolent non-Hodgkin's lymphoma. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation *7*, 680-687.
- Voso, M.T., Fenu, S., Latagliata, R., Buccisano, F., Piciocchi, A., Aloe-Spiriti, M.A., Breccia, M., Criscuolo, M., Andriani, A., Mancini, S.*, et al.* (2013). Revised International Prognostic Scoring System (IPSS) predicts survival and leukemic evolution of myelodysplastic syndromes significantly better than IPSS and WHO Prognostic Scoring System: validation by the Gruppo Romano Mielodisplasie Italian Regional Database. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *31*, 2671-2677.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.
- Walter, M.J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M.*, et al.* (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia *25*, 1153-1158.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873-878.
- Wang, J., Fernald, A.A., Anastasi, J., Le Beau, M.M., and Qian, Z. (2010a). Haploinsufficiency of Apc leads to ineffective hematopoiesis. Blood *115*, 3481-3488.
- Wang, J., Kong, G., Liu, Y., Du, J., Chang, Y.I., Tey, S.R., Zhang, X., Ranheim, E.A., Saba-El-Leil, M.K., Meloche, S.*, et al.* (2013). Nras(G12D/+) promotes leukemogenesis by aberrantly regulating hematopoietic stem cell functions. Blood *121*, 5203-5207.
- Wang, J., and Pantopoulos, K. (2011). Regulation of cellular iron metabolism. Biochem J *434*, 365-381.

- Wang, K., Li, M., and Hakonarson, H. (2010b). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164.
- Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D.S., Zhang, L.*, et al.* (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine *365*, 2497-2506.
- Ward, A.J., and Cooper, T.A. (2010). The pathobiology of splicing. J Pathol *220*, 152-163.
- Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Coller, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., Perl, A.E.*, et al.* (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. Cancer cell *17*, 225-234.
- Wardrop, D., and Steensma, D.P. (2009). Is refractory anaemia with ring sideroblasts and thrombocytosis (RARS-T) a necessary or useful diagnostic category? British journal of haematology *144*, 809-817.
- Watanabe-Okochi, N., Kitaura, J., Ono, R., Harada, H., Harada, Y., Komeno, Y., Nakajima, H., Nosaka, T., Inaba, T., and Kitamura, T. (2008). AML1 mutations induced MDS and MDS/AML in a mouse BMT model. Blood *111*, 4297-4308.
- Welch, J.S., Ley, T.J., Link, D.C., Miller, C.A., Larson, D.E., Koboldt, D.C., Wartman, L.D., Lamprecht, T.L., Liu, F., Xia, J.*, et al.* (2012). The origin and evolution of mutations in acute myeloid leukemia. Cell *150*, 264-278.
- Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.
- Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Luhrmann, R., and Query, C.C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. The EMBO journal *20*, 4536-4546.
- Wingert, R.A., Galloway, J.L., Barut, B., Foott, H., Fraenkel, P., Axe, J.L., Weber, G.J., Dooley, K., Davidson, A.J., Schmid, B.*, et al.* (2005). Deficiency of glutaredoxin 5 reveals Fe-S clusters are required for vertebrate haem synthesis. Nature *436*, 1035-1039.
- Woll, P.S., Kjallquist, U., Chowdhury, O., Doolittle, H., Wedge, D.C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q.*, et al.* (2014). Myelodysplastic Syndromes Are Propagated by Rare and Distinct Human Cancer Stem Cells In Vivo. Cancer cell.
- Wong, J.C., Zhang, Y., Lieuw, K.H., Tran, M.T., Forgo, E., Weinfurtner, K., Alzamora, P., Kogan, S.C., Akagi, K., Wolff, L.*, et al.* (2010). Use of chromosome engineering to model a segmental deletion of chromosome band 7q22 found in myeloid malignancies. Blood *115*, 4524-4532.
- Wu, S.J., Kuo, Y.Y., Hou, H.A., Li, L.Y., Tseng, M.H., Huang, C.F., Lee, F.Y., Liu, M.C., Liu, C.W., Lin, C.T.*, et al.* (2012). The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. Blood.
- Xing, Y., Xu, Y., Chen, Y., Jeffrey, P.D., Chao, Y., Lin, Z., Li, Z., Strack, S., Stock, J.B., and Shi, Y. (2006). Structure of protein phosphatase 2A core enzyme bound to tumor-inducing toxins. Cell *127*, 341-353.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T.*, et al.* (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer cell *19*, 17-30.
- Xu, X., Hou, Y., Yin, X., Bao, L., Tang, A., Song, L., Li, F., Tsang, S., Wu, K., Wu, H.*, et al.* (2012). Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell *148*, 886-895.
- Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A.*, et al.* (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature *467*, 1114-1117.

- Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J.*, et al.* (2009). IDH1 and IDH2 mutations in gliomas. The New England journal of medicine *360*, 765-773.
- Ye, H., Jeong, S.Y., Ghosh, M.C., Kovtunovych, G., Silvestri, L., Ortillo, D., Uchida, N., Tisdale, J., Camaschella, C., and Rouault, T.A. (2010). Glutaredoxin 5 deficiency causes sideroblastic anemia by specifically impairing heme biosynthesis and depleting cytosolic iron in human erythroblasts. The Journal of clinical investigation *120*, 1749-1761.
- Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865-2871.
- Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M.*, et al.* (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. FEBS J *278*, 4870-4880.
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M.*, et al.* (2011a). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature.
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M.*, et al.* (2011b). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69.
- Zebisch, A., and Sill, H. (2008). Are mouthwashes a reliable source of constitutional DNA in patients with leukemia? Leukemia research *32*, 1164-1165.
- Zhang, M., Atkinson, R.L., and Rosen, J.M. (2010). Selective targeting of radiation-resistant tumor-initiating cells. Proceedings of the National Academy of Sciences of the United States of America *107*, 3522-3527.
- Zheng, L., and Dean, D.R. (1994). Catalytic formation of a nitrogenase iron-sulfur cluster. The Journal of biological chemistry *269*, 18723-18726.

Appendix